ROBUST ESTIMATION FOR STRUCTURAL TIME SERIES MODELS

Tan Hwee Kwan

A Thesis submitted for a degree of

Doctor of Philosophy

The London School of Economics

Febuary 1990

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Dedicated to

Jehovah,

for His faithfulness and gracious provision

Abstract

This thesis aims at developing robust methods of estimation in order to draw valid inference from contaminated time series . We concentrate on additive and innovation outliers in structural time series models using a state space representation . The parameters of interest are the state , hyperparameters and coefficients of explanatory variables .

Three main contributions evolve from the research . Firstly , a filter named the approximate Gaussian sum filter is proposed to cope with noisy disturbances in both the transition and measurement equations . Secondly , the Kalman filter is robustified by carrying over the M–estimation of scale for i.i.d observations to time–dependent data . Thirdly , robust regression techniques are implemented to modify the generalised least squares transformation procedure to deal with explanatory variables in time series models . All the above procedures are tested against standard non–robust estimation methods for time series by means of simulations . Two real examples are also included .

Acknowledgements

Throughout the development of this research , I am indebted to my supervisor Dr M. Knott for his constant help and guidance . The many stimulating discussions with him are of great benefit and my intellectual debt to him is substantial . I would also like to thank my supervisor Prof. A.C. Harvey for his helpful advice and support . His comments and ideas have significantly affected this thesis .

Thanks are also due to my family and many friends for their prayers and encouragement . The financial support from the National University of Singapore is gratefully acknwledged . It is the loving support given and sacrifices made by my husband , Teik Yong , that has made this research possible . I wish to offer him my sincere and heartfelt thanks .

Table of Contents

List of Tables

List of Figures

Acronyms

ACF	Autocorrelation function
AGSF	Approximate Gaussian sum filter
AIC	Akaike's information criterion
AMLE	Approximate maximum likelihood estimator
AO	Additive outlier
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
GLST	Generalised least squares transformation
GRE	Gross error sensitivity
i.i.d.	Independent and identically distributed
IO	Innovation outlier
MA(1)	First order moving average
ML	Maximum likelihood
MLE	Maximum likelihood estimator
MSE	Mean square error
OLS	Ordinary least squares
RA	Residual autocovariance
TLS	Trimmed least squares

Chapter One

Introduction

1.1 The outlier problem

It is common in most statistical analysis to assume that observations are homogeneously generated from the same probability structure . However , in real data , we often find the appearance of outliers , that is , observations which seem to be inconsistent with the main group of data . Whether we consider such observations to be aberrant depends on our beliefs concerning the underlying probability model . Outliers are a sample phenomenon and we have to decide if they are genuine members of the population .

Discordant observations can be due to errors in reading , recording or calculating the data . Their appearance may also be due to some non–repetitive exogenous interventions or execution faults in assembling the data . All these give rise to outliers of deterministic nature which can be removed or replaced once they are identified . Sometimes , outliers are just manifestations of the inherent variability in the data generation process . Such discrepant observations are not controllable and are regarded as having random or inexplicable nature.

The presence of outliers violates the homogeneity assumption of data . It follows that the statistical properties of estimators which are based on this assumption are invalidated . In fact , contaminants create difficulties in our attempt to represent the population from which we believe the sample is drawn . Hence , it is necessary for us to recognize , interpret and make allowance for outlying observations with the help of appropriately designed techniques.

1.2 Treatment of outliers

There are two distinct approaches in the treatment of outliers . Firstly , we can "test" to determine whether an aberrant observation should be retained or ignored . The second approach is to use all the data but minimizing the influence of discordant observations . The choice between these two methods depends on the aim of the investigation .

1.2.1 Testing for discordance

Although we cannot be sure that a discrepant observation is spurious , we may feel that the loss in accuracy in the statistical analysis caused by rejecting several good values is small compared to the loss caused by keeping one bad value . This leads to testing outliers with the intention of throwing them away . There are several ways to detect deviant observations . For example , we can use graphical procedures or perform hypothesis testing .

Using graphical displays to exhibit outliers is , perhaps , the simplest and most popular method . In the time series context , we usually study the residual plot after fitting the model . Although there exist some graphical procedures specifically aimed at detecting outliers , see Gnanadesikan (1977) , very often , other assumptions underlying the statistical analysis are also under investigation . It is sometimes difficult to differentiate between the departures from the various assumptions . Therefore , this method should be treated as an informal screening process used only as a preliminary step in the analysis .

To introduce some degree of objectivity in the rejection of outliers , a statistical test is often used to decide whether aberrant observations belong to the main population ; see Barnett and Lewis (1977) for a detailed summary of these statistics . For linear models , most tests use residuals from least squares as a source for identifying outliers . Fox (1972) discusses the likelihood ratio test for autoregressive time series models . But what turns out to be a outlier in one test may not be considered as a outlier in another . Which result , then , should the investigator use ? Besides , we cannot , in most cases , construct tests which are globally uniformly most powerful .

As mentioned earlier , we can first apply conventional statistical procedures on our data set and then inspect the residuals for discordant values . Detecting outliers in this manner , with the help of graphical methods or hypothesis testing , gives rise to two problems . Firstly , the initial or trial fit can cause the effect of an outlier to be smeared across several residuals so that good observations are mistaken as outliers , see Bruce and Martin (1989) . Secondly , the initial results can be so badly distorted that bad observations do not appear as outliers . Aberrant data points may also appear in a particular configuration resulting in their effect being masked , see Atkinson (1986) .

Once the outliers are identified in a time series , we can throw away these aberrant observations and proceed with analyzing the remaining data set based on the original model . Otherwise , we can modify the model to incorporate the outliers in a non–discordant fashion or concentrate attention on deviant values so as to identify unexpected phenomenon in the data generation process . Relevant action is taken depending on our interest in the practical situation .

1.2.2 Robust procedures

Sometimes our interest is simply in inferring characteristics of the basic model regardless of the presence or nature of outliers . Then contaminants only have nuisance value and we wish to accommodate them . We turn to statistical procedures constructed to draw valid inference about the population from which the bulk of the sample is drawn and which will not be seriously distorted by contaminants . Such methods are said to be robust against the presence of outliers because they minimize the impact of bad data points . Here , outliers themselves are not of prime concern and we can estimate or test parameters of the basic model in spite of them .

Unless we are sure of the underlying distribution of the data set , which is rare when dealing with real data , estimators are no longer derived from some optimal principles for a particular distributional assumption . Rather , robust estimators are derived to achieve good performance over a broad class of distributions , especially long–tailed ones . The presence of rogue values can be modelled by letting the underlying distribution be a mixture of the distributions of "good" and "bad" data , the latter having a substantially bigger variance than the former . By considering robustness in terms of distributional assumptions , we leave aside problems involving dependence among observations . By using heavy–tailed distributions , the robust estimators will be less vulnerable to the effect of outliers . Huber (1977) stated that "for most practical purposes , 'distributionally robust' and 'outlier resistant' are interchangable" . The concept of robustness is well discussed in Hampel (1971) (1974) . A good review of robust estimators is given by Huber (1972) (1977) . In the time series context , Abraham and Box (1979) ; Denby and Martin (1979) ; Dejongh and Dewet (1985) ; and Martin (1980) propose some robust estimation techniques . Martin and Yohai (1985) gives an informative review on the robust estimators for

autoregressive moving average models .

1.3 Outliers in Time Series

The analysis of time series is an important area of statistics that has many practical applications . Because the observations are not independent , the effect of outliers now depends on their position in the series and their time configuration , i.e. , whether they are isolated or occur in patches . In fact , any discrepant observation tends to influence adjacent values due to the correlation pattern of the basic process . It becomes more difficult to detect bad data points as they need not be extreme values and can be cloaked to some extent by the general structure of the process .

1.3.1 Outlier models in time series

In due consideration of the above , care has to be taken in modeling outliers in time series . Martin and Yohai (1986) proposed the following general replacement model for a series of length T .

Let y_t be the contaminated process ;

x_t the core process which is often Gaussian ;

ω_t the contamination process ;

and z_t^γ a 0–1 process where $0 \leq \gamma \leq 1$ and $p(z_t^\gamma = 1) = \gamma$.

Then the general replacement model is given by

$$y_t = (1 - z_t^{\gamma})x_t + z_t^{\gamma}\omega_t \qquad (1.3.1)$$

for t=1,2,...,T

The probability that any one observation is free from contamination being $1-\gamma$.

This general replacement model allows for two distinct types of outliers in time series , namely , additive and innovation outliers.

(a) Additive outlier model

An additive outlier is caused by an isolated measurement or execution error superimposed on an otherwise reasonable realization of the process . It is not reflected in the values of adjacent observations . The manifestation of such outliers can sometimes be dramatic and obvious .

An additive outlier model is obtain by letting ω_t and x_t in the general replacement model have the following relationship .

$$\omega_t = x_t + v_t \qquad (1.3.2a)$$

The contaminated process is , thus , given by

$$y_t = x_t + z_t^{\gamma}v_t \qquad (1.3.2b)$$

for t=1,2,...,T

The time configuration of abnormal records can be modelled by the definition of z_t^{γ}.

(b) Innovation Outlier Model

This is a more inherent form of contamination , eg , it can be caused by a local change in the mean or variance of the series . Innovation outliers are reflected through the correlation structure of the process in neighboring values . Hence , detection is more difficult than additive outliers .

The difference between additive outlier and innovation outlier models can more clearly be seen by considering the pth order autoregressive process :

$$x_t = \sum_{k=1}^{p} \rho_k x_{t-k} + \epsilon_t \qquad (1.3.3)$$

for t=1,2,...,T

The additive outlier model is

$$y_t = x_t + z_t^{\gamma} v_t \qquad (1.3.4a)$$

$$x_t = \sum_{k=1}^{p} \rho_k x_{t-k} + \epsilon_t \qquad (1.4.4b)$$

where ϵ_t follows a Gaussian process.

Additive outliers affect only the measurement equation but not the actual autoregressive process .

The innovation outlier model has

$$y_t = x_t \tag{1.3.5a}$$

$$x_t = \sum_{k=1}^{p} \rho_k x_{t-k} + \epsilon_t \tag{1.3.5b}$$

where ϵ_t has a long–tailed distribution .

Here , there is no measurement error but any shock to the system influences both current and subsequent observations.

1.4 Time series models

1.4.1 ARIMA modeling

The robustified version of some statistical procedures has been developed for stationary time series , see Deny and Martin (1979) , Fox (1972) and Martin (1980) . Autoregressive moving average (ARMA) models are frequently used in representing stationary time series . Non–stationary series are first differenced to achieve stationarity and then modeled in this way . This class of models is known as the autoregressive integrated moving average (ARIMA) models , see Box and Jenkins (1976) . Robustness work done in this context can be found in DeJongh and DeWet (1985) , Martin et al (1983) , Martin and Yohai (1985) (1986) . However , such an approach in dealing with the problem of outliers in time series poses several difficulties .

Firstly , differencing the series results in more contaminated time points . Suppose we have a set of data , say ,

$$y_1, y_2, \cdots, \overset{*}{y_n}, \cdots, y_T$$

with only one error occurring at time n . Taking first difference , i.e. ,

$$\Delta y_t = y_t - y_{t-1}$$

gives us

$$\Delta y_2, \Delta y_3, \cdots, \Delta \overset{*}{y_n}, \Delta \overset{*}{y_{n+1}}, \cdots, \Delta y_T$$

Notice that the single aberrant observation affects the differenced series at two different time points , namely , at n and n+1 . Such proliferation of aberrant points accentuates the outlier problem when we deal with differenced series .

Besides , a single abnormal record can obscure the model fitting process . The model identification stage of ARIMA modeling usually involves the use of the autocorrelation function (ACF) . When we estimate the ACF , we weight deviated time series by their absolute distance from the series mean . The presence of an outlier will cause both an increase in the series variance as well as a drop in the covariance leading to an underestimation of the low lags of the ACF . Thus , due to the nature of the estimated ACF , a wrong model could be fitted . Furthermore , contaminated data often show certain nonstationary characteristics which makes model identification problematic .

To overcome these and other difficulties , we shall consider the outlier problem within another framework of time series modeling .

1.4.2 Structural time series models

In structural time series modeling , observations y_t are regarded as arising from a composition of independent processes . Each component has a direct interpretation in terms of the structure of the model although it is not directly observable . For instance , the "basic structural model" , see Harvey and Peters (1984) takes the form

$$y_t = m_t + s_t + \epsilon_t \qquad (1.4.1)$$

where

m_t is the trend which represents the long term movement of the series ;

s_t is the seasonal component which takes into account seasonal effects by means of a fixed period periodic function ;

and ϵ_t is the irregular term which captures the random and temporary effects on the series.

The process generating the trend can be regarded as giving a local approximation to the linear trend , i.e. ,

$$m_t = m_{t-1} + b_{t-1} + \eta_t \qquad (1.4.2a)$$

$$b_t = b_{t-1} + u_t \qquad (1.4.2b)$$

where η_t and u_t are distributed independently of each other and over time with mean zero and variances σ^2_η and σ^2_u respectively . The level and slope both change slowly over time according to a random walk process .

The process generating the seasonal component is

$$s_t = -\sum_{j=1}^{n-1} s_{t-j} + \omega_t \qquad (1.4.3)$$

where ω_t is an independently distributed disturbance term with mean zero and variance σ_ω^2 and n is the number of seasons in a year . The seasonal pattern is , therefore , slowly changing but by a mechanism which ensures that the sum of the seasonal components over any n consecutive time periods has expected value zero and a variance which remains constant over time .

Let us consider the simple random walk plus noise model which is also known as the local level model . It is a series with only trend and irregular component , the former following a random walk .

$$y_t = m_t + \epsilon_t \qquad (1.4.4a)$$

$$m_t = m_{t-1} + \eta_t \qquad (1.4.4b)$$

where ϵ_t and η_t are assumed to be white noise . A white noise variable has zero mean and a constant variance . Taken as a sequence , they are uncorrelated .

Suppose there is a recording mistake at time n which only affects the observation at that time . Then at time n ,

$$y_n = m_n + \epsilon_n + v_n \qquad (1.4.5a)$$

and $$m_n = m_{n-1} + \eta_n \qquad (1.4.5b)$$

This corresponds to the additive outlier model . However , if a step change occurs at time n , a more permanent effect is produced and the model becomes

$$y_n = m_n + \epsilon_n \tag{1.4.6a}$$

and $$m_n = m_{n-1} + v_n + \eta_n \tag{1.4.6b}$$

This corresponds to the innovation outlier model.

1.4.3 State space form and the Kalman filter

Structural time series models can easily be put in state space form . First , we have a set of variables , called the state vector $\underset{\sim}{x}_t$, which changes over time . This state vector is related to the observations via the measurement equation

$$y_t = \underset{\sim}{z}_t'\underset{\sim}{x}_t + \epsilon_t \tag{1.4.7a}$$

and is generated by a transition equation

$$\underset{\sim}{x}_t = T_t\underset{\sim}{x}_{t-1} + \underset{\sim}{\eta}_t \tag{1.4.7b}$$

Both ϵ_t and $\underset{\sim}{\eta}_t$ are assumed to be white noise with variance σ_ϵ^2 and variance-covariance respectively . These two equations make up the state space representation . Notice that the local level model is already in the state space form with state variable m_t .

The state space form for the basic structural model in equations (1.4.1) to (1.4.3) is given by

$$y_t = (1\ 0\ 1\ 0\ \ldots\ 0)\underset{\sim}{x}_t + \epsilon_t \tag{1.4.8a}$$

$$\begin{bmatrix} m_t \\ b_t \\ s_t \\ s_{t-1} \\ \cdot \\ s_{t-n+2} \end{bmatrix} = \left[\begin{array}{cc|ccc|c} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & -1 & -1 & \cdots & -1 \\ \hline 0 & 0 & & & & \\ \cdot & \cdot & & I_{n-2} & & \underset{\sim}{0} \\ 0 & 0 & & & & \end{array}\right] \begin{bmatrix} m_{t-1} \\ b_{t-1} \\ s_{t-1} \\ s_{t-2} \\ \cdot \\ s_{t-n+1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ u_t \\ \omega_t \\ 0 \\ \cdot \\ 0 \end{bmatrix}$$

(1.4.8b)

$$\underset{\sim}{x}_t = \begin{bmatrix} m_t \\ b_t \\ s_t \\ s_{t-1} \\ . \\ s_{t-n+2} \end{bmatrix} = \left[\begin{array}{cc|cccc} 1 & 1 & 0 & 0 & . . . & 0 \\ 0 & 1 & 0 & 0 & . . . & 0 \\ 0 & 0 & -1 & -1 & . . . & -1 \\ . & . & 1 & 0 & . . . & 0 \\ . & . & . & . & . . . & . \\ 0 & 0 & 0 & 0 & . . . & 0 \end{array}\right] \begin{bmatrix} m_{t-1} \\ b_{t-1} \\ s_{t-1} \\ s_{t-2} \\ . \\ s_{t-n+1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ u_t \\ \omega_t \\ 0 \\ . \\ 0 \end{bmatrix} \tag{1.4.8b}$$

Additive outliers , being spurious values occurring in the observed series , affect only the observation equation . On the other hand , innovation outliers corresponding to rogue values that affect the driving noise , have a more lasting effect on subsequent values of the series . Innovation outliers disturb the state equation only .

Once the model is put in state space form , we can apply the Kalman filter for sequential estimation of the state vector . The Kalman filter comprises prediction and updating equations for the state vector and its covariance matrix .

Under the Gaussian assumption of the disturbances , the prediction equations give the best one–step ahead estimate of the state vector , $\underset{\sim}{a}_{t/t-1}$ and its mean–square error matrix , $P_{t/t-1}$. The prediction equations are :

$$\underset{\sim}{a}_{t/t-1} = T_t \underset{\sim}{a}_{t-1} \tag{1.4.9a}$$

$$P_{t/t-1} = T_t P_{t-1} T_t + Q \tag{1.4.9b}$$

where $\underset{\sim}{a}_{t-1}$ and P_{t-1} are the best estimate of the state vector and its covariance matrix at time t–1 .

Whenever a new observation y_t is available , the new information is

incorporated to give an optimal estimate of the state at time t , $\underset{\sim}{a}_t$. The updating equations are

$$\underset{\sim}{a}_t = \underset{\sim}{a}_{t/t-1} + P_{t/t-1}\underset{\sim}{z}_t(y_t - \underset{\sim}{z}_t'\underset{\sim}{a}_{t/t-1})/f_t \qquad (1.4.10a)$$

and $$P_t = P_{t/t-1} - P_{t/t-1}\underset{\sim}{z}_t\underset{\sim}{z}_t'P_{t/t-1}/f_t \qquad (1.4.10b)$$

with $$f_t = \underset{\sim}{z}_t'P_{t/t-1}\underset{\sim}{z}_t + \sigma_\epsilon^2 \qquad (1.4.10c)$$

where f_t is the prediction error variance at time t . The Kalman filter produces optimal estimates under the normality assumption of the disturbances ϵ_t and η_t . The filter also gives the prediction errors and their associated variances which can be used in the evaluation of the log–likelihood for the observations . This is useful when we want to evaluate the maximum likelihood estimate of the parameters in the model such as the variance of the disturbance terms . Prediction of future values beyond the series can also be made using the filter .

1.5 Robust estimation in time series

As mentioned earlier , additive and innovation outliers affect the measurement and transition equations respectively . These discrepant observations can be incorporated into the model by letting the disturbance terms ϵ_t or $\underset{\sim}{\eta}_t$, as appropriate , take on long–tailed distributions . These are symmetric distributions with more mass on the tails than the corresponding Gaussian . Such a heavy–tailed

distribution gives rise to occasional large values . An example of this kind of distribution is the contaminated normal distribution . This is a mixture of normal densities given by :

$$(1-\gamma)\, N(0,\sigma_1^2) \; + \; \gamma\, N(0,\sigma_2^2)$$

where σ_2^2 is much bigger than σ_1^2 and γ is small . This model arises when the disturbances are assumed to be normal with variance σ_1^2 but a fraction γ of them are subjected to gross errors . The contaminating distribution is normal with a big variance σ_2^2 . Letting the noise terms ϵ_t and $\underset{\sim}{\eta}_t$ follow the contaminated normal distribution implies that the normality assumption of the disturbances are no longer valid . It follows that the one step ahead estimates of the state , $\underset{\sim}{a}_{t/t-1}$, from the Kalman filter are no longer optimal but they are still the best linear unbiased estimates (BLUE) . Unfortunately , such linear least squares estimates can behave quite badly in the presence of outliers , see Huber (1972) .

To illustrate the bad behaviour , a time series of length fifty is generated from a local level model . The measurement noise term follows a contaminated normal distribution with γ=5% , σ_1^2=1 and σ_2^2=100 while the system noise is normally distributed , N(0,1) . This is to allow for additive outliers only in the series . A plot of this series is found in figure 1.1 . Running the Kalman filter on the data set produces one–step ahead predictions which are also plotted in figure 1.1 . The one–step ahead predictions are badly distorted by aberrant values and each outlier affects several subsequent predictions . On the other hand , corresponding values from a robust filter are less sensitive to the influences of heavy–tailed distributions , see figure 1.1 .

In the next chapter , we shall look at robust filtering or robust sequential estimation in time series . Various filters have been proposed by Masreliez (1975) , Martin and Masreliez (1977) , Ershov and Liptser (1978) , West (1981) ,Martin et al (1983) and Guttman and Pena (1985) to produce robust estimates of the state given past observations . These filters are computationally attractive because they have the same recursive structure as the Kalman filter . Modifications are carried out at the updating equations in order to bound observations with large prediction errors . Such filters are based on Masreliez assumption which assumes normality for the state prediction density . We show in the chapter that this assumption is not valid in the presence of extreme additive or innovation outliers . This means that the above filters cannot cope with such a situation .

Non–Gaussian filters which are derived directly from the recursive relations underlying the sequential estimation process , namely the Gaussian sum filter proposed by Alspach and Sorenson (1972) , Kitagawa's filter (1988) and Student–t distribution filter (1989) are discussed . The latter uses mixtures of Student–t distributions to represent the density functions while Kitagawa's filter uses a numerical method based on piecewise linear approximations of the density functions . On the other hand , the Gaussian sum filter uses mixtures of normal terms to represent each density function . Although these filters are designed to deal with series having both forms of outliers present , they are computationally inconvenient.

A modification of the Gaussian sum filter is described in Chapter three . The basic procedure is to collapse the normal terms in each mixture in order to curb the exponential growth in the number of components in each density . This method of collapsing is similar to that used by Harrison and Stevens (1976) . The resulting approximate Gaussian sum filter (AGSF) is computationally feasible and can handle

situations where both additive and innovation outliers are present . This filter is developed in the context of the local level model but can be easily extended to a multi–state system . Parameter estimation is carried out by the maximum likelihood estimation procedure whilst the level of contamination can be determined by a broad grid search . If we have a rough idea of the level of contamination , a robustified version of the likelihood can be used for parameter estimation . Application of the AGSF to a real data set is illustrated .

Chapter four deals with the scale estimation problem , i.e. , we shift our attention from the sequential estimation of the state to the estimation of the hyperparameters $\sigma^2_{1\epsilon}$ and $\sigma^2_{1\eta}$. In the presence of additive outliers , the Kalman filter tends to overestimate the hyperparameter $\sigma^2_{1\epsilon}$. This is because the estimate includes the variation of the contaminating component so that we are actually estimating the overall variance instead of $\sigma^2_{1\epsilon}$. After examining the scale model for steady state observations , we carry over the M–estimation technique from i.i.d. situation to time series context . The multidimensional Huber function , see Hampel (1986) , is employed to bound the influence of outliers . A sampling procedure for calibrating the estimators in order to achieve Fisher consistency is described . The M–estimation of the hyperparameters is carried out using the Kalman filter and is robust against additive outliers only .

In the final chapter , the local level model is extended to include explanatory variables . The generalised least squares transformation (GLST) method , see Ansley and Kohn (1985) , is robustified . This modified procedure involves an iteration between the M–estimation of scale and trimmed least squares estimation of the regression coefficients . Robust estimates can be obtained for both the regression parameters and the hyperparameters as the explanatory variables take on several

different models . An application of the robustified GLST procedure to a real data set is discussed .

Key

——————— y_t

— — — — $a_{t/t-1}$ from Kitagawa's filter

– – – – – – $a_{t/t-1}$ from Kalman filter

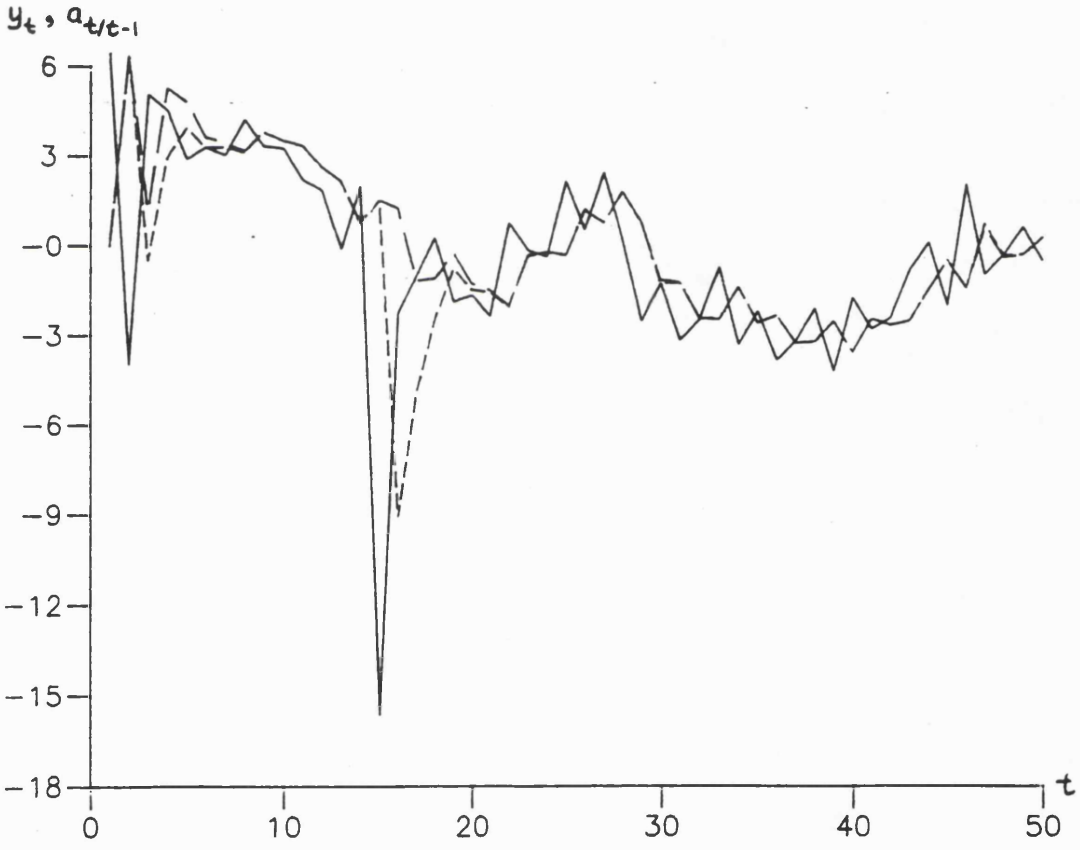


Figure 1.1 Plot of series with additive outliers and one–step ahead predictions from Kitagawa's filter and the Kalman filter

Chapter Two

A review of robust filters

2.1 Recursive Relations

2.1.1 The Model

In the previous chapter , see equation (1.4.7) the state space representation of a univariate time series is given by

$$y_t = \underset{\sim}{z}_t' \underset{\sim}{\alpha}_t + \epsilon_t \qquad (2.1.1a)$$

$$\underset{\sim}{\alpha}_t = T_t \underset{\sim}{\alpha}_{t-1} + \underset{\sim}{\eta}_t \qquad (2.1.1b)$$

for t = 1,2,...,T

To simplify the problem , let us first consider a time invariant system whose state consists of only one element . The measurement equation becomes

$$y_t = z\alpha_t + \epsilon_t \qquad (2.1.2a)$$

and the state evolves according to

$$\alpha_t = h\alpha_{t-1} + \eta_t \qquad (2.1.2b)$$

for t = 1,2,...,T

The variances of the disturbance terms , σ_ϵ^2 and σ_η^2 , are known as the hyperparameters of the model . In this chapter , we are concerned with the robust estimation of the state α_t .

2.1.2 Recursive Relations using Bayes' theorem

A recursive scheme for the state α_t can be formed by applying Bayes' theorem . Specifically , we can express the state's prediction density at time t , $p(\alpha_t/y_1,y_2,\cdots,y_{t-1})$, in terms of the updating density at time t–1 , $p(\alpha_{t-1}/y_1,y_2,\cdots,y_{t-1})$. When a new observation y_t becomes available , it can be incorporated with this prediction density to give the updating density at time t , $p(\alpha_t/y_1,y_2,\cdots,y_t)$. Denoting all the past observations up to the current time , i.e. $\{y_1,y_2,\cdots,y_t\}$, as Y_t we have the following relations

(i) One Step Ahead Prediction

$$p(\alpha_t/Y_{t-1}) = \int_{-\infty}^{\infty} p(\alpha_t,\alpha_{t-1}/Y_{t-1})\, d\alpha_{t-1} \tag{2.1.3a}$$

$$= \int_{-\infty}^{\infty} p(\alpha_t/\alpha_{t-1})\, p(\alpha_{t-1}/Y_{t-1})\, d\alpha_{t-1} \tag{2.1.3b}$$

(ii) Updating

$$p(\alpha_t/Y_t) = p(\alpha_t/y_t,Y_{t-1}) \tag{2.1.4a}$$

$$= p(\alpha_t,y_t/Y_{t-1}) \,/\, p(y_{t-1}/Y_{t-1}) \tag{2.1.4b}$$

$$= p(y_t/\alpha_t)\, p(\alpha_t/Y_{t-1}) \,/\, p(y_t/Y_{t-1}) \tag{2.1.4c}$$

where the normalizing constant $p(y_t/Y_{t-1})$ is given by

$$p(y_t/Y_{t-1}) = \int_{-\infty}^{\infty} p(\alpha_t/Y_{t-1})\, p(y_t/\alpha_t)\, d\alpha_t \qquad (2.1.5)$$

We can express the initial density $p(\alpha_0/y_0)$ in the same form as in equation (2.1.4c) . That is ,

$$p(\alpha_0/y_0) = p(y_0/\alpha_0)\, p(\alpha_0) \,/\, p(y_0) \qquad (2.1.6)$$

We observe that the prediction density $p(\alpha_t/\alpha_{t-1})$ in (2.1.3b) depends on the system noise density $p(\eta_t)$ and the transition equation (2.1.2b) . On the other hand , the updating density $p(y_t/\alpha_t)$ in (2.1.4c) involves the observation noise density $p(\epsilon_t)$ and the measurement equation (2.1.2a) . Hence , a knowledge of the distribution of the disturbances , $p(\epsilon_t)$ and $p(\eta_t)$ enables us to determine the prediction and updating densities of the state at any time t . We shall call these four densities pertinent densities .

2.1.3 Pertinent Densities having Elliptical Distributions

If the initial state and disturbance terms are Gaussian , the above set of equations (2.1.3) to (2.1.5) can be evaluated . In fact , at any time t , both the prediction and updating densities are normal . Hence , it is enough to consider the first two moments of each distribution . This is why the Kalman filter evaluates only the means and variances at each stage . The filter makes use of the fact that normal distributions are closed under linear transformations and that the conditional expectations is linear in the linear conditioning variables . A class of elliptical random processes , see Chu (1973) , which includes Gaussian densities has such

properties . Therefore , when the joint distribution of the initial state and the disturbances are elliptical , the Kalman filter can still be applied . However , the random variables α_0 , ϵ_t and η_t must in general be dependent because the joint distribution of two independent elliptical random variables is not elliptical except when both marginals are Gaussian . This lack of independence amongst the variables in the non–Gaussian case , is a restrictive condition on the use of the Kalman filter.

2.1.4 Non–Gaussian Pertinent Densities

In the non–Gaussian case , the pertinent densities cannot be characterized by a finite number of moments . As a result , the system cannot be solved in a closed form and it becomes necessary to evaluate the non–normal densities explicitly at each stage . Besides , the conditional mean , which is the minimum variance estimate in the Gaussian case , is no longer a linear function of the measurement data . At the same time , the conditional variance now depends on the observations . Generally , it is not possible to evaluate analytically the integrals in equations (2.1.3) to (2.1.5) and the densities cannot be determined in most applications . This is , in particular , true when we let the densities take on heavy–tailed distributions to allow for outliers . Some form of approximation is , therefore , necessary to realize the formulae in the recursive relations .

2.2 Kitagawa's Filter

Kitagawa (1987) uses a numerical method based on piecewise linear approximations to evaluate the density functions in the recursive relations . First , he represents a non–stationary time series in the state space form as in equations (2.1.2) . Neither the observation nor the system noise is necessarily Gaussian . Rather , they can take on heavy–tailed distribution to incorporate aberrant observations . Sequential estimation of the state is performed by carrying out prediction and filtering as in section 2.1.2 .

Each density function in the recursive relations is approximated by a piecewise linear function alternatively known as a first order spline . Thus , we can specify each density by the number of segments , k ; the position of the nodes , n_0 , n_1 ,...,n_k and the value at each node . The outermost nodes n_0 and n_k are selected so that the essential domain of the density is covered . This is to ensure that the error caused by truncation is negligible . Hence , for long–tailed distributions , the extreme nodes should be located further out . The degree of accuracy we have in the approximation depends on the number of segments . Increasing the number of nodes increases the accuracy at the expense of more computation . The values at the nodes define the entire function under the piecewise linearity assumption of the function .

To simplify the computation , the same node scheme is used for all the density functions . Once we assume the distribution of the noise terms , we can calculate its density at any point . Using notations $p_t(\alpha)$, $f_t(\alpha)$, $r(\epsilon_t)$ and $q(\eta_t)$ for the prediction , updating , measurement noise and transition noise densities respectively , we have the following algorithms.

(i) One Step Ahead Prediction

For each $\alpha = n_0 , n_1 ,\ldots,n_k$

$$p_t(\alpha) = p(\alpha_t/Y_{t-1})$$

$$= \int_{-\infty}^{\infty} p(y/Y_{t-1})\, p(\alpha/y)\, dy$$

$$= \int_{-\infty}^{\infty} f_{t-1}(y)\, q(\alpha - hy)\, dy$$

$$= \sum_{i=1}^{k} \int_{n_{i-1}}^{n_i} f_{t-1}(y)\, q(\alpha - hy)\, dy \qquad (2.2.1)$$

Assuming the functions $f_{t-1}(y)$ and $q(\alpha - hy)$ to be linear in any segment $\{n_{i-1},n_i\}$, each integral can be approximated as follows

$$\int_{n_{i-1}}^{n_i} f_{t-1}(y)\, q(\alpha - hy)\, dy$$

$$\approx [f_{t-1}(n_{i-1})q(\alpha - hn_{i-1}) + f_{t-1}(n_i)q(\alpha - hn_i)]\, (n_i - n_{i-1})/2 \qquad (2.2.2)$$

Substituting (2.2.2) into (2.2.1) gives the prediction density as

$$p_t(\alpha) \approx \sum_{i=1}^{k} [f_{t-1}(n_{i-1})q(\alpha - hn_{i-1}) + f_{t-1}(n_i)q(\alpha - hn_i)]\, (n_i - n_{i-1})/2 \qquad (2.2.3)$$

(ii) Updating

For each $\alpha = n_0 , n_1 ,\ldots,n_k$

$$f_t(\alpha) = p(\alpha_t/Y_t)$$
$$= p_t(\alpha)r(y_t - z\alpha)/c \qquad (2.2.4)$$

where the normalizing constant c is given by

$$c = p(y_t/Y_{t-1})$$

$$= \int_{-\infty}^{\infty} p_t(y)\, r(y_t - zy)\, dy$$

$$= \sum_{i=1}^{k} \int_{n_{i-1}}^{n_i} p_t(y)\, r(y_t - zy)\, dy \qquad (2.2.5)$$

As before , we assume $p_t(y)$ and $r(y_t - zy)$ to be linear functions on any interval $\{n_{i-1},n_i\}$ so that we can approximate c by the following relation

$$c \approx \sum_{i=1}^{k} [p_t(n_{i-1})r(y_t - zn_{i-1}) + p_t(n_i)r(y_t - zn_i)]\, (n_i - n_{i-1})/2 \qquad (2.2.6)$$

To start the recursions , we need the density of the initial state , $p(\alpha_0)$. If no prior information on the state is available , we can use a non–informative prior or a diffuse prior , see Harvey and Peters (1984) . For instance , we can let the initial state α_0 follow a normal distribution with a zero mean and a very big but finite

variance . Given the distributions of the noise terms $r(\epsilon_t)$ and $q(\eta_t)$, Kitagawa's filter produces an approximation to the prediction and updating densities at each stage .

Estimation of the parameters in the system (2.1.2) can be easily carried out with Kitagawa's filter . Maximum likelihood estimates are obtained by maximizing the log–likelihood function which can be expressed as follows

$$\ln L(Y_T) = \sum_{t=1}^{T} \ln p(y_t/Y_{t-1}) \qquad (2.2.7)$$

This breaks the joint density of dependent observations down to a simpler form . An approximation to each conditional density $p(y_t/Y_{t-1})$ is already produced when running Kitagawa's filter , see (2.2.6) . Thus , the computation of the log–likelihood function can be easily incorporated into the filter .

We use a maximization routine from the NAG library named E04JAF together with an algorithm for Kitagawa's filter to estimate parameters in the model in equation (2.1.2) . Two function subroutines are included to evaluate the observation noise and system noise densities . Recall the time series generated at the end of Chapter one which only has additive outliers , let us specify correctly in Kitagawa's filter the distribution of the measurement and system noise terms . They follow a contaminated normal distribution with parameters $\gamma = 5\%$, $g = 100$, $\sigma_\epsilon^2 = 1$, see 3.1.2a) and a normal distribution N(0,1) respectively . Running Kitagawa's filter on this time series gives us a set of prediction density functions . The mean of each of these densities is computed and plotted in figure 1.1 . Comparing these values with those produced by the Kalman filter shows that Kitagawa's filter is robust against outliers .

Kitagawa employs a direct method for his filter but a great deal of computation is required . This is especially true when we extend the filter to analyse models with multidimensional state . In this case , the system disturbance term will have a multivariate distribution and we need to deal with the convolution of such density functions . This means that the amount of computation is greatly increased . The number of nodes can be reduced by using a variable mesh or a moving mesh , see Kitagawa (1987) . However , this makes the algorithm more complex which takes up more computation time . Higher order splines can also be used but they may not be numerically stable . In conclusion , although Kitagawa's filter is easy to understand and implement , it is not computationally attractive .

2.3 Gaussian sum filter

Alspach and Sorenson (1971) proposes another non–Gaussian filter based on the recursive relations in section 2.1.2 . They investigate efficient and simplified methods for approximating and computing the conditional probability density for the state . Each density is approximated by a convex combination of Gaussian density functions . Every Gaussian sum approximation is a valid density function and converges uniformly to any density of practical concern . We note that the long–tailed contaminated normal distribution used to model the disturbances is itself a Gaussian mixture .

The system under consideration is given in (2.1.2) . We denote the normal density for variable y with mean μ and variance σ^2 by $N_y(\mu,\sigma^2)$. Suppose the initial

state density is represented by

$$p(\alpha_0) = \sum_{i=1}^{n_0} \delta_{0i}\, N_\alpha(a_{0i}, P_{0i}) \tag{2.3.1a}$$

Let the measurement and system noise , ϵ_t and η_t , be serially and pairwise independent , with density functions taking the form

$$p(\epsilon_t) = \sum_{i=1}^{m_t} \gamma_{ti}\, N(0,\sigma^2_{\epsilon i}) \tag{2.3.1b}$$

$$p(\eta_t) = \sum_{i=1}^{s_t} \beta_{ti}\, N(0,\sigma^2_{\eta i}) \tag{2.3.1c}$$

Then , the Gaussian sum filter is made up of a group of Kalman filters with a set of mean values and corresponding variances to be computed . These means and variances are used to construct prediction and updating densities as Gaussian sum .

Suppose that the prediction density at time t is described by

$$p(\alpha_t/Y_{t-1}) = \sum_{i=1}^{l_t} \rho_{ti}\, N_\alpha(a_{t/t-1,i}, P_{t/t-1,i}) \tag{2.3.2}$$

Then , the updating density is given by

$$p(\alpha_t/Y_t) = \sum_{i=1}^{l_t} \sum_{j=1}^{m_t} c_{ij}\, N_\alpha(\mu_{ij}, \sigma^2_{ij}) \tag{2.3.3}$$

where

$$\mu_{ij} = a_{t/t-1,i} + P_{t/t-1,i}\, z\, (y_t - z a_{t/t-1,i}) / f_{ij} \tag{2.3.4a}$$

$$\sigma^2_{ij} = P_{t/t-1,i} - P^2_{t/t-1,i}\, z^2 / f_{ij} \tag{2.3.4b}$$

$$f_{ij} = P_{t/t-1,i}\, z^2 + \sigma^2_{\epsilon j} \tag{2.3.4c}$$

$$c_{ij} = \rho_{ti}\gamma_{tj}\, N_{y_t}(b_{ij}, \xi^2_{ij}) \Big/ \sum_{l=1}^{l_t} \sum_{m=1}^{m_t} \rho_{tl}\gamma_{tm}\, N_{y_t}(b_{lm}, \xi^2_{lm}) \tag{2.3.5a}$$

using the notations

$$b_{ij} = z a_{t/t-1,i} \tag{2.3.5b}$$

$$\xi^2_{ij} = z^2 P_{t/t-1,i} + \sigma^2_{\epsilon j} \tag{2.3.5c}$$

Note that $c_{ij} \geq 0$ for all i,j and

$$\sum_{i=1}^{l_t} \sum_{j=1}^{m_t} c_{ij} = 1 \tag{2.3.5d}$$

If we rewrite the updating density as

$$p(\alpha_t / Y_t) = \sum_{i=1}^{n_t} \delta_{ti}\, N_\alpha(a_{t/t,i}, P_{t/t,i}) \tag{2.3.6}$$

Then the prediction density is

$$p(\alpha_t / Y_{t-1}) = \sum_{i=1}^{n_t} \sum_{j=1}^{s_t} \delta_{ti}\beta_{tj}\, N_\alpha(h a_{t/t,i}, \lambda^2_{ij}) \tag{2.3.7a}$$

where

$$\lambda_{ij}^2 = hP_{t/t,i} + \sigma_{\eta j}^2 \qquad (2.3.7b)$$

Since the initial state density has been put in the same form as the posterior density $p(\alpha_t/Y_t)$, see (2.3.1a) , the Gaussian sums repeat themselves from one stage to the next whereby (2.3.2) and (2.3.6) are the general forms for an arbitrary stage . Hence , the Gaussian sum is almost reproducing except for the growing number of components in the mixture . It follows that the densities are not described by a fixed number of parameters . If the Gaussian sums for all the prior densities in (2.4.1) consist of only one term , that is , if they are normal variables , the filter described above in (2.3.2) – (2.3.7) , reduces to the Kalman filter . In fact , the terms μ_{ij} and σ_{ij}^2 in (2.3.4) and the terms $ha_{t/t,i}$ and λ_{ij}^2 in (2.3.7) represent the Kalman filter equations for the ij th density combination of the Kalman filters . Parameter estimation is carried out by the maximum likelihood procedure as in the Kitagawa's filter .

One disadvantage of this method is the problem of finding the appropriate Gaussian sum representation . The second difficulty is the exponential growth in the number of terms in the sums as the steps are processed . This problem can be reduced by combining terms with nearly equal moments and ignoring terms with diminishing weighting factors . These mechanisms reduce the number of terms in the sum substantially but introduce some error into the calculations . Hence , like the Kitagawa's filter , the Gaussian sum filter is computationally inefficient .

2.4 Student–t distribution filter

R.J. Meinhold and N.D. Singpurwalla (1989) approach the non–Gaussian filter problem from a Bayesian viewpoint using the recursive relations in section 2.1.2 . This filter produces robust estimates of the state in a state space model such as the one in equation (2.1.2) by letting the disturbance terms and the initial state assume independent Student–t distributions . This will result in the prediction and updating densities following "poly–t" distributions with no closed form representation . A recursive approximation scheme is proposed , implementing two theorems on the convergence of the pertinent densities depending on their degrees of freedom .

Suppose the updating density at time t–1 is represented by a mixture of Student–t distributions , that is ,

$$\alpha_{t-1} \sim \sum_{j=1}^{N_{t-1}} \delta_{j,t-1}\, t(a_{j,t-1}, P_{j,t-1}, n) \qquad (2.4.1a)$$

$$\text{with} \sum_{j=1}^{N_{t-1}} \delta_{j,t-1} = 1 \qquad (2.4.1b)$$

where $a_{j,t-1}$, $P_{j,t-1}$ and n denote the mean, variance and degrees of freedom respectively.
Like the Gaussian sum filter , the prediction density at time t is formed by a componentwise convolution of each term in the mixture with the system noise density. Hence , the prediction density is given by

$$\alpha_{t/t-1} \sim \sum_{j=1}^{N_{t-1}} \delta_{j,t-1}\, t(ha_{j,t-1}, h^2 P_{j,t-1} + \sigma_\eta^2, n) \tag{2.4.2}$$

We note that the number of components in the sum remains the same since the disturbance term has only one component in its density .

According to equation (2.1.4) , each of the N_{t-1} terms in the updating density at time t is given by

$$\frac{t(ha_{j,t-1}, h^2 P_{j,t-1} + \sigma_\eta^2, n)\, t(y_t, \sigma_\epsilon^2, m)}{\int_{-\infty}^{\infty} t(ha_{j,t-1}, h^2 P_{j,t-1} + \sigma_\eta^2, n)\, t(y_t, \sigma_\epsilon^2, m)\, dx} \tag{2.4.3}$$

If this component is unimodal , it will be approximated by a Student–t density with n degrees of freedom ; centered at $a_{j,t}$ which is set to the mode of the posterior ; weight $\delta_{j,t}$ remains the same at $\delta_{j,t-1}$; and scale $P_{j,t}$ is determined by setting the approximating density equal to the height of the actual density at the mode .

However , if the term in (2.4.3) is bimodal , then it will be approximated by a mixture of two Student–t densities , each with n degrees of freedom . These densities will be centered at the two modes of the original densities . The scale parameters are found by equating the curvature of the approximating densities to the curvature of the actual density at their modes . The weights of the mixture are

$$\delta_{j1,t} = \frac{\delta_{j,\,t-1}\; t(y_t, \sigma_\epsilon^2, m)}{t(y_t, \sigma_\epsilon^2, m) + t(ha_{j,t-1}, h^2 P_{j,t-1} + \sigma_\eta^2, n)} \tag{2.4.4a}$$

$$\delta_{j2,t} = \delta_{j,t-1} - \delta_{j1,t} \tag{2.2.4b}$$

Hence , the number of terms in this mixture of Student–t distribution also increases as we process the series . Another difficulty associated with this filter is the determination of the degrees of freedom for the distributions of the disturbance terms.

2.5 Masreliez Assumption and the gain function

2.5.1 Masreliez Assumption

We shall now turn our attention to non–Gaussian filters which retain the computationally attractive recursive structure of the Kalman filter . These filters have been derived based on the Masreliez (1975) assumption of a normal distribution for the state prediction density , $p(\alpha_t/Y_{t-1})$, at each step . It is because of this assumption that the filters can be put in the Kalman filter form . As discussed in section 2.1.3 , we know that the Masreliez assumption is true when there is no contamination in the series . In practice , this assumption is usually closely satisfied when there are only additive outliers present in the series . However , the presence of extreme additive or innovation outliers invalidates the assumption .

We generate a series of length one hundred from the local level model with the measurement noise assuming a contaminated noise distribution and the system noise following a normal distribution . Additive outliers are modelled by making the variance of the second component fifty times bigger than the first , see pg 32 . By running Kitagawa's filter on this data set with only additive outliers , we obtain a sequence of state prediction densities . A normal probability plot of the state

prediction density is made at an outlier point and its adjacent point . These plots can be found in figure 2.1 . We observe that the normal plot at both the outlier and non–outlier point is a straight line , indicating the normality of the state prediction densities at such points . Thus , when the series has only additive outliers , the Masreliez assumption is closely satisfied .

Next we generate another series from the local level model but with the system noise following a contaminated normal distribution . We allow for innovation outliers in the series by letting the variance of its second component be fifty times bigger than the first . Running Kitagawa's filter on this data set produces yet another sequence of state prediction densities . The normal probability plots here , see figure 2.2 , show a definite deviation from normality at both outlier as well as non–outlier points . It follows that the Masreliez assumption is not valid when innovation outliers are present in the series .

2.5.2 Kalman filter and the gain function

Attempts have been made to modify the Kalman filter to handle a non–Gaussian observation density . Consider the Kalman filter equations for system in (2.1.2) . They are

(a) Prediction

$$a_{t/t-1} = h\, a_{t-1} \tag{2.5.1a}$$

$$P_{t/t-1} = h^2 P_{t-1} + \sigma_\eta^2 \tag{2.5.1b}$$

(b) Updating

$$a_t = a_{t/t-1} + P_{t/t-1}\, z\, \frac{v_t}{f_t} \tag{2.5.2a}$$

$$P_t = P_{t/t-1} - P^2_{t/t-1}\, z^2\, \frac{1}{f_t} \tag{2.5.2b}$$

where

$$v_t = y_t - z\, a_{t/t-1} \tag{2.5.3a}$$

$$f_t = z^2\, P_{t/t-1} + \sigma^2_\epsilon \tag{2.5.3b}$$

The terms v_t and f_t are the one–step ahead prediction error , also known as innovation , and its variance respectively . Alternatively , we can express (2.5.2a) and (2.5.2b) as

$$a_t = a_{t/t-1} + P_{t/t-1}\, z\, \psi\left(\frac{v_t}{\sqrt{f_t}}\right)\frac{1}{\sqrt{f_t}} \tag{2.5.4a}$$

$$P_t = P_{t/t-1} - P^2_{t/t-1}\, z^2\, \psi'\left(\frac{v_t}{\sqrt{f_t}}\right)\frac{1}{f_t} \tag{2.5.4b}$$

where the psi–function ψ is an identity function , i.e. ,

$$\psi(u) = u \tag{2.5.4c}$$

and ψ' is the derivative of ψ .

A plot of this psi–function is given in figure 2.3a .

The psi–function is actually a gain function operating on scaled residuals , $\frac{v_t}{\sqrt{f_t}}$.

Hence , its shape determines the influence of these residuals on the updated estimate . In the case of the Kalman filter , residuals of different magnitude are given equal emphasis since the psi–function is a straight line . Robustification of this filter can be done by selecting a gain function which de–emphasizes large residuals usually caused by outliers . This is similar to finding a robust estimate by using the influence curve , see Hampel (1974) . Consider a series from the linear system (2.1.2) with a small probability that any observation is an additive outlier . Based on Masreliez' assumption , a robust filter takes on the same prediction (2.5.1) and updating equations (2.5.4a)–(2.5.4b) as the Kalman filter . However , to discount aberrant observations , different gain functions are used . This allows the psi–function to act as a non–linear transformation on the residuals to desensitize the procedure to outliers.

2.6 Robust filters based on Masreliez assumption

2.6.1 Missing value filter

Martin et al (1983) apply the hard rejection rule for the gain function , i.e. ,

$$\psi(u) = \begin{cases} u & \|u\| \le b \\ 0 & \|u\| > b \end{cases} \qquad (2.6.1)$$

A plot of this function is found in figure 2.3b . Whenever the innovations are greater than b times its standard deviation , the current observation is discarded and no

updating occurs . This is equivalent to identifying observations with big residuals and then treating them as missing values . The choice of b depends on the compromise between robustness and efficiency of the filter . If b is kept small , more observations will be regarded as outliers . Thus , while the procedure is well protected from discordant observations , it becomes less efficient when more good observations are thrown away . Usually , b takes the value 2 or 3 .

Outliers can be identified using the scaled innovations only if the variances of the noise terms , σ_ϵ^2 and σ_η^2 are known . Otherwise , an initial robust estimate for the scale of the innovations , denoted by $\hat{f}_t$, is needed . In which case , the criterion for modification of the psi–function is replaced by $\|v_t\| > b\ \hat{f}_t$. The maximum likelihood procedure can , then , be applied to estimate the parameters in the model . This also applies to the rest of the filters in this section .

2.6.2 Minimax filter

Another robust filter proposed by Martin and Masreliez (1977) generates min–max estimates , $a_{t/t-1}$, with respect to the least favourable distribution , F_0 , see Huber (1964) . Under Masreliez' assumption , the prediction equations turn out to be the same as those in (2.5.1) . However , updating equation for variance (2.5.4b) is approximately given by

$$P_t \approx P_{t/t-1} - P_{t/t-1}^2\, z^2\, E_F\, \psi'\left(\frac{v_t}{\sqrt{f_t}}\right)\frac{1}{f_t} \qquad (2.6.2)$$

where ψ' is the first derivative of the gain function ψ ,
and the expectation is taken over the distribution F .

The updating equation for the state still takes the form (2.5.4a) with the gain function , see figure 2.3c , being defined as

$$\psi(u) = \begin{cases} u & \|u\| \leq b \\ b \ \mathrm{sign}(u) & \|u\| > b \end{cases} \qquad (2.6.3)$$

where the choice of b is affected by the same consideration as before . This means that all innovations greater than b times the scale of prediction error will have the same impact on the updated estimate . In this way , the procedure is desensitized towards random large disturbances . Accompanying this choice of psi–function , the least favourable distribution , F_0 , is Gaussian in the middle with exponential tails . Maximum likelihood estimates can be obtained from this filter as in the previous one .

For strict inequality to hold in (2.6.2) , another gain function is applied . The corresponding estimates produced are known as p–value estimates , see Martin and Masreliez (1975) and the psi–function is defined as

$$\psi(u) = \begin{cases} \tan(\frac{1}{2sb}) \,/\, sb & \|u\| \leq b \\ \tan(\frac{1}{sb}) \ \mathrm{sgn}(u) \,/\, sb & \|u\| > b \end{cases} \qquad (2.6.4)$$

See figure 2.3d for a plot of this psi–function . The least favourable distribution , F_0 , here has a density which goes like $\cos^2(u)$ in the middle with exponential tails . Considering only observations whose scaled residuals are smaller than b , more emphasis is given to those with bigger innovations . This is appropriate only when the distributions of the disturbances are less heavy–tailed than the corresponding Gaussian density . Therefore , the use of p–value estimates is not recommended

when we want to deal with outliers . If the psi–function (2.6.4) is used , the expression the right of (2.6.2) is actually an upper bound for P_t which may be approached arbitrarily .

2.6.3 Masreliez filter

Masreliez (1975) suggests a filter that provides protection against outliers associated with heavy–tailed deviations from Gaussianness and which reduces to the Kalman filter in the Gaussian case . Under the assumption of normal state prediction density , the filter is derived using the recursive relations in section 2.1.2 . The prediction and updating equations remain as (2.5.2) and (2.5.4a)–(2.5.4b) respectively . But now the gain function is the negative score function , i.e. ,

$$\psi\left(\frac{v_t}{\sqrt{f_t}}\right) = -\frac{\partial\, p(y_t/Y_{t-1})}{\partial y_t} \Big/ \left[p(y_t/Y_{t-1})/\sqrt{f_t}\right] \qquad (2.6.5)$$

The shape of this psi–function clearly depends on $p(y_t/Y_{t-1})$ which is the convolution of the state prediction density and the measurement noise density . When the density is Gaussian , the filter reduces to the Kalman filter . To cope with outliers , $p(y_t/Y_{t-1})$ takes on a symmetric long–tailed distribution leading to a odd symmetric non–linear psi–function which de–emphasizes the influence from large residuals , see figure 2.3e .

Parameter estimation can be carried out by direct maximization of the log–likelihood function (2.2.7) noting that the density $p(y_t/Y_{t-1})$ is already computed at each stage . Now , the convolution of a normal density , $p(x_t/Y_{t-1})$

with a contaminated normal density , $p(\epsilon_t)$ can easily be evaluated . But if the observation noise term follows another long–tailed distribution , the convolution may be difficult to evaluate . In any case , the distribution of the observation noise term is assumed to be known .

2.6.4 West's filter

To avoid this need for the convolution of the densities , West (1981) rewrites the updating equations in (2.5.4) as

$$a_t = a_{t/t-1} + P_t g(v_t) \tag{2.6.6a}$$

$$P_t^{-1} = \left[P_{t-1} + \sigma_\eta^2\right]^{-1} + G(v_t) \tag{2.6.6b}$$

where

$$g(u) = -\frac{\partial \ln p_\epsilon(u)}{\partial u} \tag{2.6.6c}$$

$$G(u) = -\frac{\partial^2 \ln p_\epsilon(u)}{\partial u^2} \tag{2.6.6d}$$

and leaving the prediction equations as (2.5.1) . In this way , the gain function is now expressed in terms of the measurement noise density instead of $p(y_t/Y_{t-1})$. For certain choices of likelihood , however , g(u) redescends to zero and G(u) becomes negative for large values of u . It is , therefore , necessary to truncate G(u) , setting it to zero outside the range of positive values . This will introduce some error into the calculations .

2.6.5 Ershov and Liptser's filter

A non–optimal , non–linear filter is presented in Ershov and Liptser (1978) to cope with contamination in the measurement process . The system noise assumes a normal distribution $N(0,\sigma_\eta)$ in the linear system (2.1.2) . Recalling the general replacement model in Chapter one , the observation noise term is specified as

$$\epsilon_t \sim (1-z_t)\epsilon_t^* + z_t\omega_t \qquad (2.6.7a)$$

where

$\epsilon_t^* \sim N(0,\sigma_1^2)$

$\omega_t \sim N(0,\sigma_2^2)$ and

z_t is 0–1 process with $p(z_t=1) = \gamma$

such that

$$\gamma\sigma_2^2 \gg (1-\gamma)\sigma_1^2 \qquad (2.6.7b)$$

As with the above filters , the prediction equations are given in (2.5.1) . Updating , however , takes the following form

$$a_t = a_{t/t-1} + P_{t/t-1}\, z\, \frac{v_t}{f_t^*} \qquad (2.6.8a)$$

$$P_t = P_{t/t-1} - P_{t/t-1}^2\, z^2\, \frac{1}{f_t^*} \qquad (2.6.8b)$$

$$f_t^* = (1-s_t)\sigma_1^2 + s_t\sigma_2^2 + P_{t/t-1} \qquad (2.6.8c)$$

where s_t is an estimate of z_t . It is a indicator function given by

$$s_t = I\left\{ \|v_t\| \geq c(t,\alpha) \right\} \qquad (2.6.9a)$$

where the constant c is chosen to satisfy

$$P(\|v_t\| \geq c(t,\alpha) \,/\, z_t = 0) = \alpha \qquad (2.6.9b)$$

The term α is a pre–assigned number .

Unlike the previous filters , robustification is carried out through the modification of the Kalman gain , $\frac{P_{t/t-1}}{f_t}$, instead of the gain function $\psi(u)$. Suppose that the estimate s_t is replaced by the mean of the z_t process which is γ in the above equations , a linear filter is obtained . Simulation results in the paper show that the non–linear filter performs better that the linear one .

2.6.6 Guttman and Pena's filter

Guttman and Pena (1985) also considered the situation whereby the measurement noise term ϵ_t follows a contaminated normal distribution while the Gaussian assumption is maintained for the driving noise term η_t . We see in section 2.3 that using the Gaussian sum filter will result in a proliferation of normal terms in both the prediction and updating densities . This can be curbed by approximating each Gaussian sum by a single normal distribution . The collapsed distribution is uniquely determined by the first and second moments of the mixture distribution .

The method of collapsing is similar to that used in Harrison and Stevens (1976) . Consider approximating a contaminated normal distribution by a single Gaussian component . That is , representing $(1-\gamma)N(\mu_1,\sigma_1^2) + \gamma N(\mu_2,\sigma_2^2)$ by

$N(\mu,\sigma^2)$. Then , the parameters μ and σ^2 are evaluated as

$$\mu = (1-\gamma)\mu_1 + \gamma\mu_2 \qquad (2.6.10a)$$

$$\sigma^2 = (1-\gamma)(\sigma_1^2 + (\mu-\mu_1)^2) + \gamma(\sigma_2^2 + (\mu-\mu_2)^2) \qquad (2.6.10b)$$

We see that the overall mean of the distribution is kept constant in (2.6.10a) . The variance of the collapsed distribution is that of the mixture plus a term which accounts for the distances between the old and new means .

In the case where innovation outliers are absent , the Gaussian sum filter can be modified to have Kalman filter's prediction equations (2.5.1) . Updating is carried out as before , see section 2.3 . At each stage , the updating density will have two normal components which are collapsed into one term in the manner described above. An algorithm for this procedure is found in Guttman and Pena (1985) .

2.7 Summary

We see in figure 2.2 that the Masreliez assumption on the normal distribution of the state prediction density is not always satisfied . It follows that the filters discussed in section 2.6 are incapable of coping with series having extreme outliers . On the other hand , non–Gaussian filters designed to deal with both additive and innovation outliers in the series , see section 2.2 and 2.4 , are difficult to implement . Therefore , an approximation to the Gaussian sum filter is proposed in the next

chapter . This filter does not depend on Masreliez assumption and is computationally attractive .

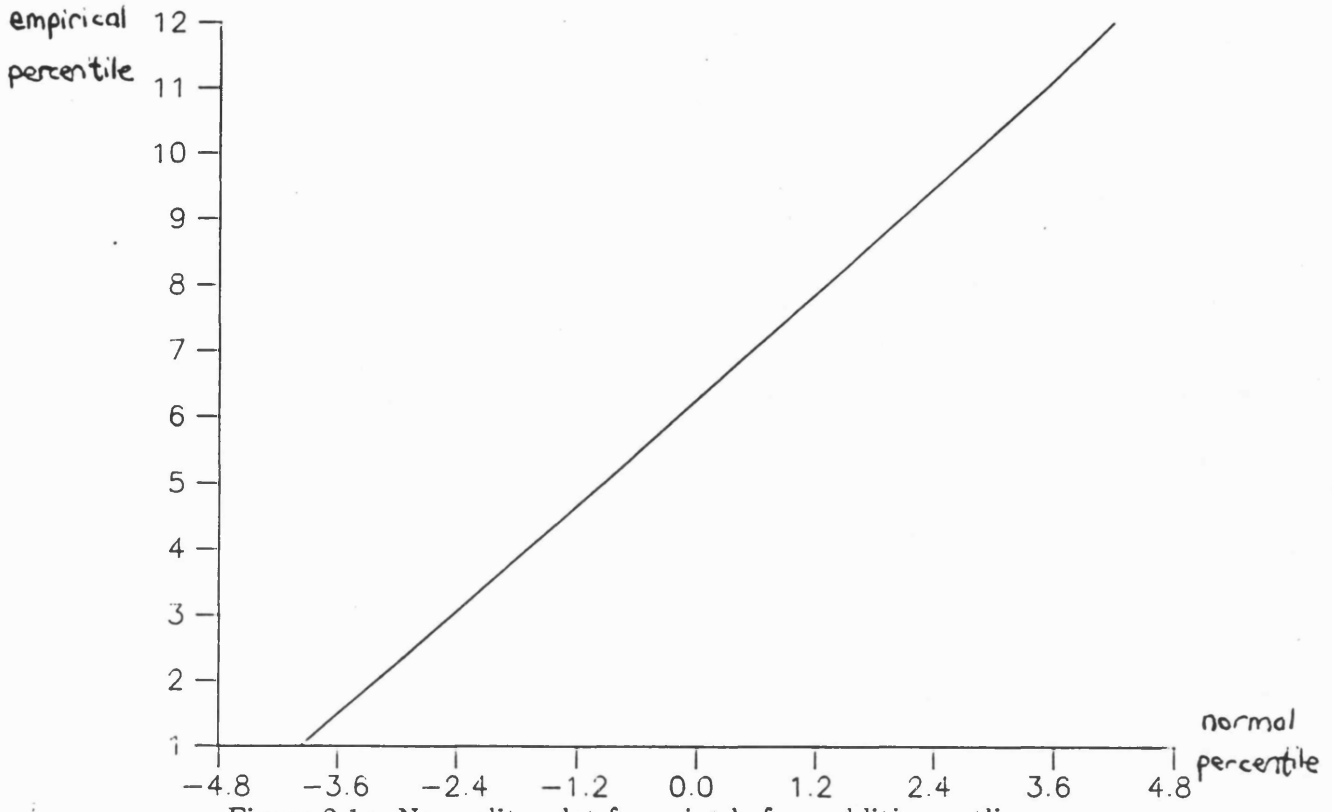


Figure 2.1a Normality plot for point before additive outlier

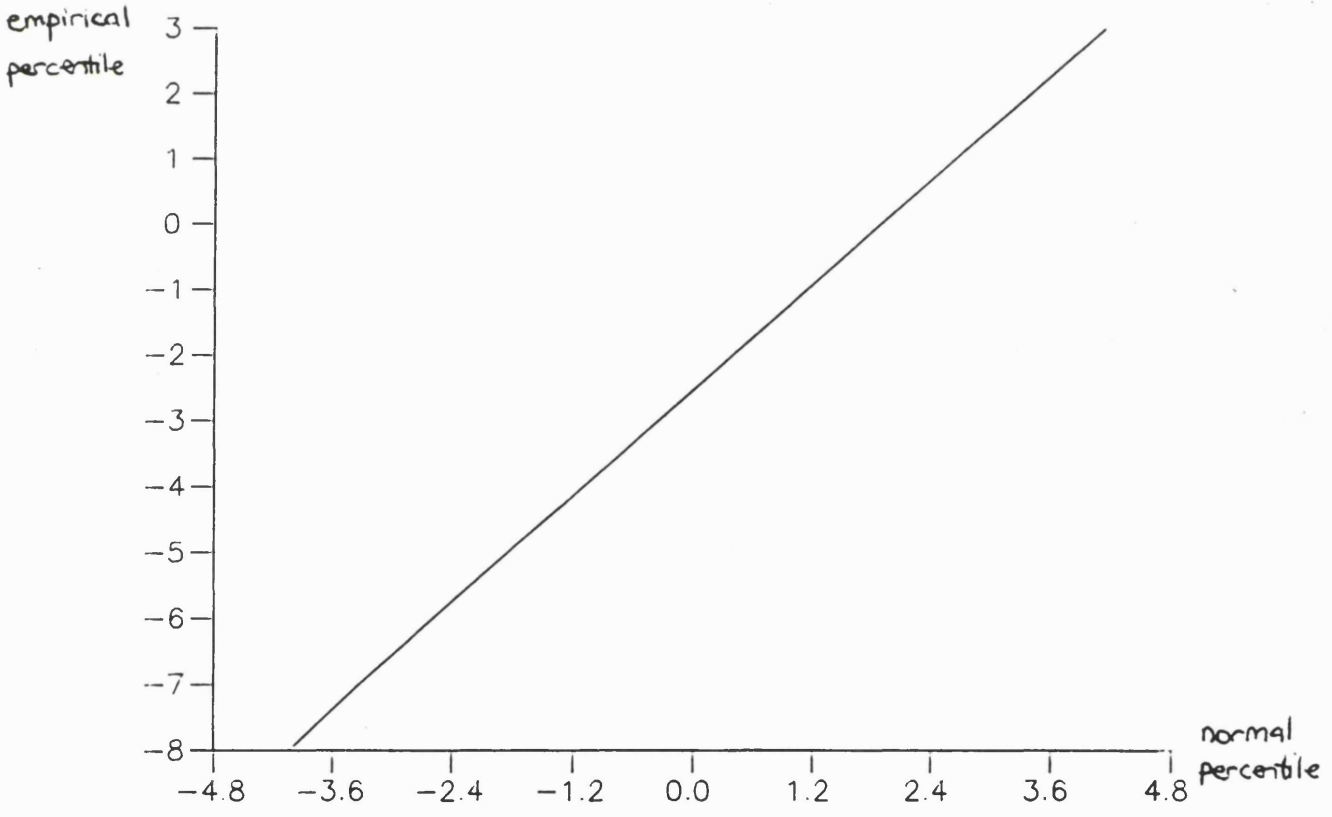


Figure 2.1b Normality plot for point at additive outlier

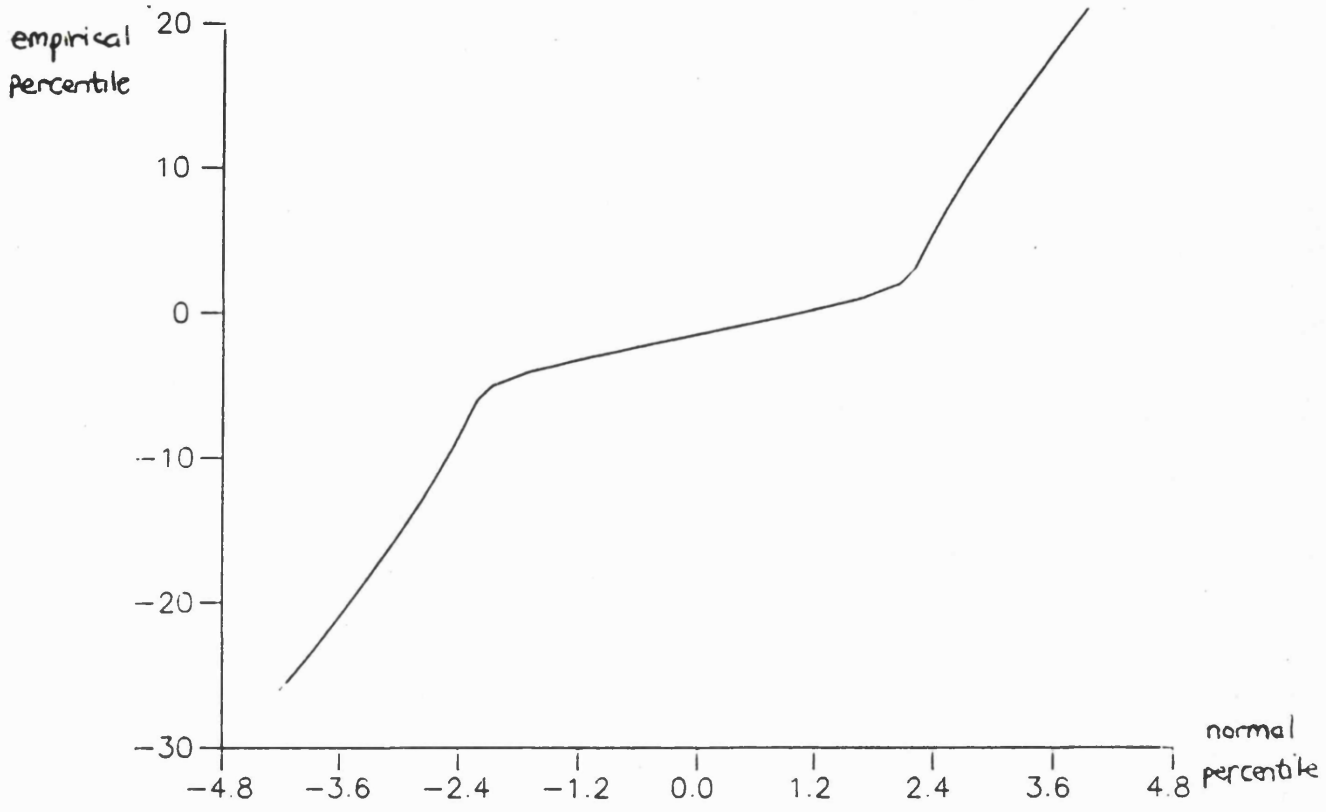


Figure 2.2a Normality plot for point before innovation outlier

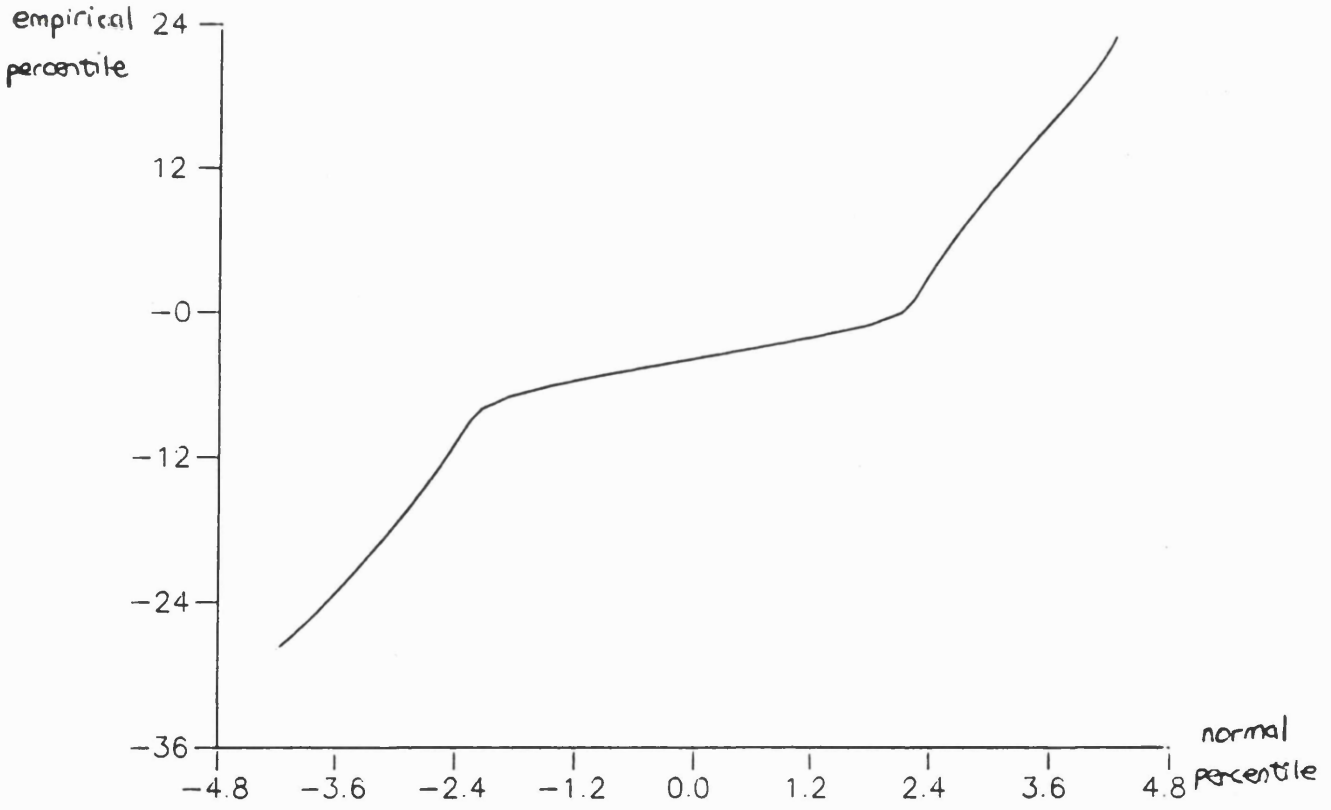


Figure 2.2b Normality plot for point at innovation outlier

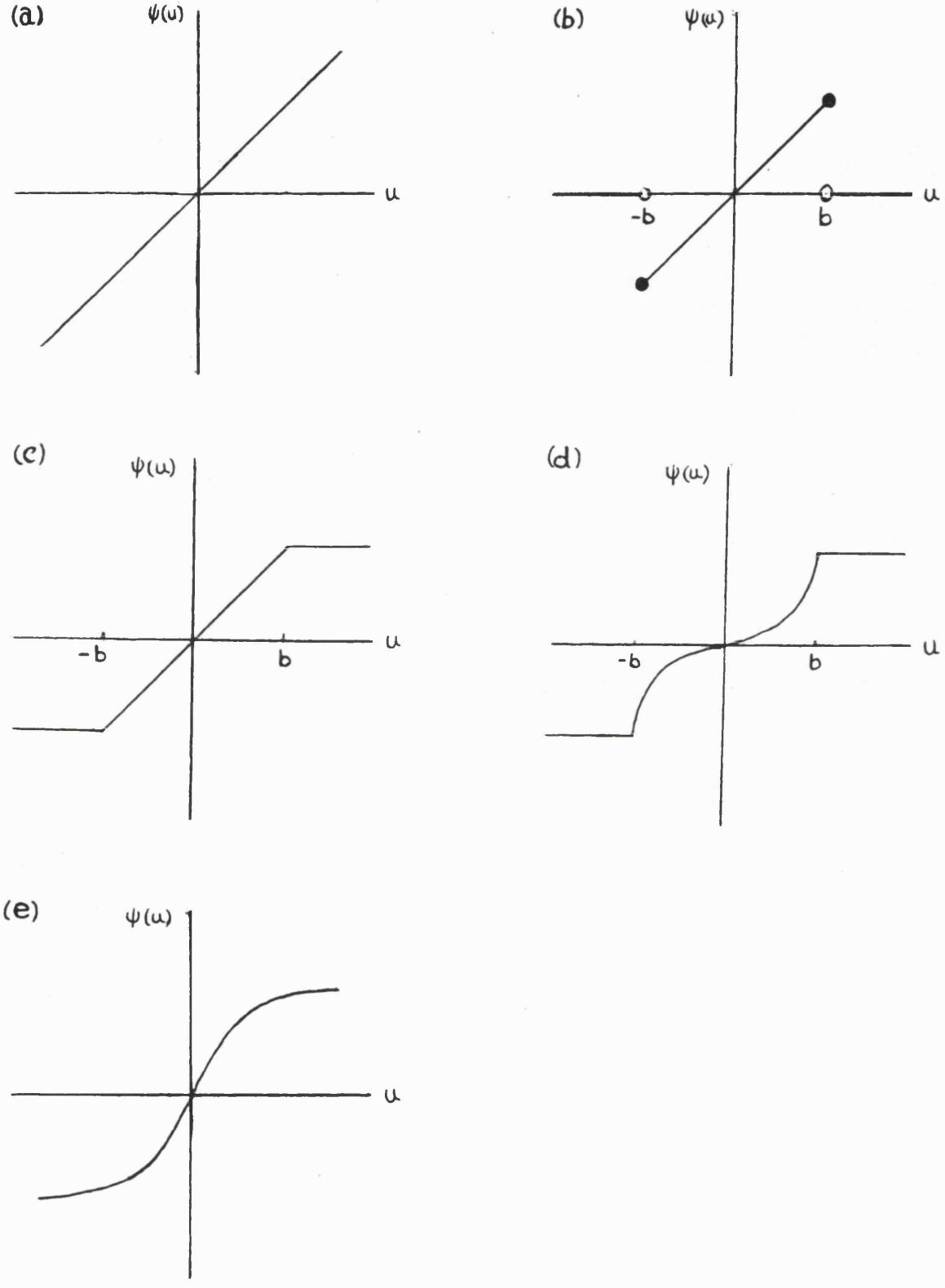


Figure 2.3 Psi–functions for (a) Kalman filter ; (b) Missing value filter ; (c) and (d) Minimax filter ; and (e) Masreliez filter

Chapter Three

The Approximate Gaussian Sum Filter (AGSF)

3.1 Introduction

3.1.1 The Gaussian Sum filter

The Gaussian sum filter , proposed by Alspach and Sorenson (1971) , is described in Chapter two . When the initial state α_0 as well as the disturbances ϵ_t and η_t of the random walk plus noise model are mixtures of normal densities , this filter produces minimum variance estimates for the state . If outliers in the model can be accommodated by letting the noise terms take on contaminated normal distributions , then the Gaussian sum filter will make a good robust filter . However , the implementation of this filter is cumbersome because it requires an exponentially increasing memory storage . Some form of approximation can be carried out to reduce the number of terms at each stage . For the resulting filter to work properly , the approximation has to be accurate at both outlier and non–outlier points . This chapter discusses a modification of the Gaussian sum filter which is easy to implement .

3.1.2 The model

Consider a linear system given by

$$y_t = z\,\alpha_t + \epsilon_t \qquad (3.1.1a)$$

$$\alpha_t = h\,\alpha_{t-1} + \eta_t \qquad (3.1.1b)$$

for $t = 1,2,...,T$

where z and h are known constants and $\|h\| < 1$. To allow for both additive and innovation outliers in the series , we let both the observation noise and system noise terms have contaminated normal distributions . That is ,

$$p(\epsilon_t) = (1-\gamma)\, N(0,\sigma_\epsilon^2) + \gamma\, N(0,g\sigma_\epsilon^2) \tag{3.1.2a}$$

$$p(\eta_t) = (1-\beta)\, N(0,\sigma_\eta^2) + \beta\, N(0,b\sigma_\eta^2) \tag{3.1.2b}$$

where $0 < \gamma,\beta < .5$; $g\sigma_\epsilon^2 \gg \sigma_\epsilon^2$ and $b\sigma_\eta^2 \gg \sigma_\eta^2$.

When a diffuse prior is employed , the initial state α_0 takes a normal distribution $N(0,\kappa)$ where $\kappa \to \infty$. However , we want to put it in the same form as (3.1.2) for ease of exposition afterwards . Hence , we write

$$p(\alpha_0) = (1-\delta_0)\, N(0,\kappa) + \delta_0\, N(0,\kappa) \tag{3.1.3}$$

3.2 Robust filtering by AGSF

3.2.1 Prediction equations

Suppose the updating density of the state at time t is a normal mixture consisting of only two terms . That is ,

$$p(\alpha_t/Y_t) = (1-\delta_t)N(a_{t/t,1},P_{t/t,1}) + \delta_t N(a_{t/t,2},P_{t/t,2}) \tag{3.2.1}$$

From equation (2.1.3) , the next prediction density is then defined as a convolution of this updating density with the system noise density . Hence , we have

$$p(\alpha_{t+1}/Y_t) = (1-\delta_t)(1-\beta)\, N(ha_{t/t,1}\,,\, h^2P_{t/t,1}+\sigma_\eta^2)$$

$$+ (1-\delta_t)\,\beta\, N(ha_{t/t,1}\,,\, h^2P_{t/t,1} + b\sigma_\eta^2)$$

$$+ \delta_t\,(1-\beta)\, N(ha_{t/t,2}\,,\, h^2P_{t/t,2} + \sigma_\eta^2)$$

$$+ \delta_t\,\beta\, N(ha_{t/t,2}\,,\, h^2P_{t/t,2} + b\sigma_\eta^2) \qquad (3.2.2)$$

The resulting prediction density comprises four normal terms which can be approximated by a Gaussian sum of only two terms . This is carried out by a collapsing method which improves that of Guttman and Pena (1985) which is illustrated in Chapter two . The procedure maintains the overall mean while computing the new variance as the old but taking into account the distance between the old and new means . It is desirable to combine two components only when they are close enough . Usually , the first term of (3.2.2) is combined with the third and the second with the fourth . However , due to the nature of the weights in the prediction density (3.2.2) , this will always lead to the collapsed terms having the same means . Hence , the collapsed prediction density will always be symmetric and such an approximation is bad at extreme outlier points .

In order to achieve a good representation of the prediction densities , it is necessary to distinguish between symmetric and asymmetric distributions . The presence of extreme additive or innovation outliers causes the state prediction densities to become skewed and/or bimodal at some points , see Kitagawa (1987) . It

is vital for the collapsed prediction density to retain such features whenever they occur . To achieve this , it is useful to note that a combination of two close components results in a unimodal symmetric function while a bimodal or heavily–skewed function is formed by two terms with considerable different means , see figure 3.1 . No distinction is made between unimodal or symmetric distributions from bimodal or skewed ones in Guttman and Pena (1985) .

The method of collapsing begins with the testing of symmetry and unimodality of the state prediction density . Referring to equation (3.2.2) , the density function is symmetric and unimodal if its two means , $a_{t/t,1}$ and $a_{t/t,2}$, are close to each other . Then , the approximate density is formed by combining the first with the third term and the second with the fourth term , see figure 3.2b . This collapsed density has a contaminated normal structure in that the first term dominates the second . On the other hand , when the distance between the two means , $a_{t/t,1}$ and $a_{t/t,2}$, is large , the prediction density is asymmetric or bimodal . In this case , the two components of the collapsed density should have greatly differing means to maintain the bimodal or skewed structure . This is achieved by combining the first two terms and then the last two terms together , see figure 3.2a . In this way , each four term normal mixture is well represented by a two term Gaussian sum .

We now have

$$p(\alpha_{t+1}/Y_t) = (1-\rho_{t+1})\, N(a_{t+1/t,1} , P_{t+1/t,1}) + \rho_{t+1}\, N(a_{t+1/t,2} , P_{t+1/t,2}) \qquad (3.2.3a)$$

Let $d_{ij} = (a_{t/t,i} - a_{t+1/t,j})h$ (3.2.3b)

We define the weights , means and variances of the above density for two separate cases below .

(a) Symmetric density

(i) weights

$$\rho_{t+1} = \beta \tag{3.2.4a}$$

(ii) means

$$a_{t+1/t,1} = (1-\delta_t)\, ha_{t/t,1} + \delta_t\, ha_{t/t,2} \tag{3.2.4b}$$

$$a_{t+1/t,2} = a_{t+1/t,1} \tag{3.2.4c}$$

(iii) variances

$$P_{t+1/t,1} = (1-\delta_t)\,(h^2 P_{t/t,1} + \sigma_\eta^2 + d_{11}^2) + \delta_t\,(h^2 P_{t/t,2} + \sigma_\eta^2 + d_{21}^2) \tag{3.2.4d}$$

$$P_{t+1/t,2} = (1-\delta_t)\,(h^2 P_{t/t,1} + b\sigma_\eta^2 + d_{12}^2) + \delta_t\,(h^2 P_{t/t,2} + b\sigma_\eta^2 + d_{22}^2) \tag{3.2.4e}$$

(b) Asymmetric density

(i) weights

$$\rho_{t+1} = \delta_t \tag{3.2.5a}$$

(ii) means

$$a_{t+1/t,1} = ha_{t/t,1} \tag{3.2.5b}$$

$$a_{t+1/t,2} = ha_{t/t,2} \tag{3.2.5c}$$

(iii) variances

$$P_{t+1/t,1} = (1-\beta)\,(h^2 P_{t/t,1} + \sigma_\eta^2) + \beta\,(h^2 P_{t/t,1} + b\sigma_\eta^2) \tag{3.2.5d}$$

$$P_{t+1/t,2} = (1-\beta)\,(h^2 P_{t/t,2} + \sigma_\eta^2) + \beta\,(h^2 P_{t/t,2} + b\sigma_\eta^2) \tag{3.2.5e}$$

The distinction between cases (a) and (b) is determined by the following rule . We say that the prediction density is symmetric iff

$$\|a_{t/t,1} - a_{t/t,2}\| \leq 3P_{t/t,1} \tag{3.2.6}$$

3.2.2 Updating equations

Consider the prediction density , equation (3.2.3a) , at time t , the state is updated when a new observation y_t becomes available . From equations (2.1.4) and (2.1.5) , we see that the updating density is written as a ratio of the product of the prediction density and the observation noise density to the convolution of these two densities . Thus , we have ,

$$p(\alpha_t/Y_t) = \sum_{i=1}^{2} \sum_{j=1}^{2} c_{ij}\, N(\mu_{ij}, \sigma_{ij}^2)/T \tag{3.2.7a}$$

where

$$c_{ij} = \rho_i \gamma_j \, N(za_{t/t-1,i} \, , \, z^2 P_{t/t-1,i} + r_j)/T \qquad (3.2.7b)$$

using the following notation

$$\rho_1 = \rho \; ; \; \rho_2 = 1-\rho$$

$$\gamma_1 = \gamma \; ; \; \gamma_2 = 1-\gamma$$

$$r_1 = \sigma_\epsilon^2 \; ; \; r_2 = g\sigma_\epsilon^2$$

$$T = \sum_{i=1}^{2} \sum_{j=1}^{2} c_{ij} \, , \text{ which is the density } p(y_t/Y_{t-1})$$

$$\mu_{ij} = a_{t/t-1,i} + P_{t/t-1,i} \, z(y_t - za_{t/t-1,i})/(z^2 P_{t/t-1,i} + r_j) \qquad (3.2.7b)$$

$$\sigma_{ij}^2 = P_{t/t-1,i} - z^2 P_{t/t-1,i}^2/(z^2 P_{t/t-1,i} + r_j) \qquad (3.2.7c)$$

Some properties of the normal density functions , see Alspach and Sorenson (1971) , are used to derive these expressions . These properties will be discussed in the next section 3.2.4 when an extension of the AGSF is considered .

Unlike the prediction density (3.2.2a) , the means of the four components in the updating density (3.2.7a) are all different . Hence , the combination of the first with the third term and the second with the fourth term does not always result in a symmetric collapsed density . The updating density is now approximated as (3.2.1)

where

(i) weights

$$\delta_t = c_{12} + c_{22} \tag{3.2.8a}$$

(ii) means

$$a_{t/t,1} = c_{11}\mu_{11} + c_{21}\mu_{21} \tag{3.2.8b}$$

$$a_{t/t,2} = c_{12}\mu_{12} + c_{22}\mu_{22} \tag{3.2.8c}$$

(iii) variances

denoting $\mu_{ij} - a_{t/t,j}$ by Δ_{ij},

$$P_{t/t,1} = c_{11}(\sigma_{11}^2+\Delta_{11}^2) + c_{21}(\sigma_{21}^2+\Delta_{21}^2) \tag{3.2.8d}$$

$$P_{t/t,2} = c_{12}(\sigma_{12}^2+\Delta_{12}^2) + c_{22}(\sigma_{22}^2+\Delta_{22}^2) \tag{3.2.8e}$$

We see in equation (3.1.3) that the prior density can be expressed as a two–term normal mixture . It follows that if the above procedure is carried out from the beginning , the prediction and updating densities at each stage are represented by Gaussian sums with only two components . The exponential growth in the number of terms in the sum is , therefore , curtailed and the amount of computation is greatly reduced . This modified filter is simple to implement even though it does not depend on Masreliez assumption . Hence , it can be applied to time series where both types of outliers are present .

3.2.3 Two examples

To compare the AGSF with the Kalman filter , a data set of length fifty is first generated from the local level model with additive outliers only . With reference to the linear system (3.1.1) and (3.1.2) , the parameters are set as follows : $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$, $\gamma = \beta = 5\%$, g = 100 , b = 1 . Assuming knowledge of these parameter values, we run the AGSF and the Kalman filter on this series . A plot of the generated state α_t and its estimate $a_{t/t-1}$ from the filters are found in figure 3.3 . It is clear from the graph that the AGSF is more robust than the Kalman filter since the former's estimate of the state are less sensitive to the outliers . Another series of length fifty is generated from the local level model but this time with both additive and innovation outliers . The parameters are kept the same as above except for the factors of contamination which are now g = 100 and b = 10 . Figure 3.4 is a plot of the state and its estimate from the AGSF and the Kalman filter . Again we see the sequential estimates of the state from the AGSF are less affected by the additive outliers compared with those from the Kalman filter .

Since a robust filter is not significantly affected by the additive outliers , it is able to track the state better . We can try to identify the aberrant observations by examining the scaled innovations . It may be possible to find explanations for these abnormalities in the series in terms of identifiable events . Appropriate adjustment may then be made to the system to deal with the outliers . The Kalman filter can be used to handle the adjusted series .

3.2.4 Extension to higher order state system

In the derivation of the Gaussian sum filter , see Alspach and Sorenson (1971), two results regarding the properties of the normal density function are used . They are

(i) For $z \neq 0$,

$$N_y(zx,\sigma^2) = N_x(\frac{y}{z}, \frac{\sigma^2}{z^2}) / z \tag{3.2.9}$$

and

(ii) $$N_x(\mu_i,\sigma_i^2)\, N_x(\mu_j,\sigma_j^2) = N_{\mu_i}(\mu_j,\sigma_i^2+\sigma_j^2)\, N_x(\mu_{ij},\sigma_{ij}^2) \tag{3.2.10a}$$

where

$$\mu_{ij} = (\mu_i\sigma_j^2 + \mu_j\sigma_i^2) / (\sigma_i^2 + \sigma_j^2) \tag{3.2.10b}$$

$$\sigma_{ij}^2 = \sigma_i^2\, \sigma_j^2 / (\sigma_i^2 + \sigma_j^2) \tag{3.2.10c}$$

The Gaussian sum expressions for the prediction and updating densities are based on these results . When analysing univariate time series having multidimensional state , see (1.4.7) , an extension of these results is necessary . Corresponding results for the multivariate case and their derivation are given as follows .

Lemma one

$$N_y(\underset{\sim}{z}'\underset{\sim}{x}\,,\, \sigma^2) = N_{\underset{\sim}{x}}(\underset{\sim}{a}y\,,\, \sigma^2(\underset{\sim}{z}\underset{\sim}{z}')^{-1}) / \sqrt{\|\underset{\sim}{z}\underset{\sim}{z}'\|} \tag{3.2.11a}$$

where

$$\underset{\sim}{z}'\underset{\sim}{a} = 1 \tag{3.2.11b}$$

Proof

Letting $K = 1 / \sqrt{2\pi\sigma^2}$, we have for the L.H.S.

$$K \exp\left\{-\frac{(y-\underset{\sim}{z}'\underset{\sim}{x})^2}{2\sigma^2}\right\}$$

$$= K \exp\left\{-\frac{(y-\underset{\sim}{z}'\underset{\sim}{x})(y-\underset{\sim}{z}'\underset{\sim}{x})}{2\sigma^2}\right\}$$

$$= K \exp\left\{-\frac{(\underset{\sim}{z}'\underset{\sim}{x}-y)(\underset{\sim}{z}'\underset{\sim}{x}-y)}{2\sigma^2}\right\}$$

$$= K \exp\left\{-\frac{[\underset{\sim}{z}'(\underset{\sim}{x}-\underset{\sim}{a}y)]\,'\,[\underset{\sim}{z}'(\underset{\sim}{x}-\underset{\sim}{a}y)]}{2\sigma^2}\right\} \quad \text{since } \underset{\sim}{z}'\underset{\sim}{a} = 1$$

$$= \frac{1}{\sqrt{\|\underset{\sim}{z}\underset{\sim}{z}'\|}} \frac{1}{\sqrt{2\pi\sigma^2 \left\|\frac{1}{\underset{\sim}{z}\underset{\sim}{z}'}\right\|}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{a}y)'\left[\frac{\underset{\sim}{z}\underset{\sim}{z}'}{\sigma^2}\right](\underset{\sim}{x}-\underset{\sim}{a}y)\right\}$$

$$= \text{R.H.S.} \qquad \text{(Q.E.D.)}$$

Lemma two

$$N_{\underset{\sim}{x}}(\underset{\sim}{\mu}_1,\Sigma_1)\, N_{\underset{\sim}{x}}(\underset{\sim}{\mu}_2,\Sigma_2) = N_{\underset{\sim}{\mu}_1}(\underset{\sim}{\mu}_2,\Sigma_1+\Sigma_2)\, N_{\underset{\sim}{x}}(\underset{\sim}{\mu}_{12},\Sigma_{12}) \qquad (3.2.12a)$$

where

$$\underset{\sim}{\mu}_{12} = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2) \qquad (3.2.12b)$$

$$\Sigma_{12} = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1} \qquad (3.2.12c)$$

Proof

(i) L.H.S. has determinant $\|\Sigma_1\|^{1/2} \|\Sigma_2\|^{1/2}$

R.H.S. has determinant $\|\Sigma_1 + \Sigma_2\|^{1/2} \|(\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\|^{1/2}$

Squaring both determinants and equating them gives

$$\|\Sigma_1 + \Sigma_2\| = \|\Sigma_1\Sigma_2 (\Sigma_1^{-1} + \Sigma_2^{-1})\|$$

$$= \|\Sigma_1 + \Sigma_2\| \qquad \text{(Q.E.D.)}$$

(ii) Consider the exponent of the L.H.S. multiplied by a factor of -2

$$(\underset{\sim}{x}-\underset{\sim}{\mu}_1)' \Sigma_1^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu}_1) + (\underset{\sim}{x}-\underset{\sim}{\mu}_2)' \Sigma_2^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu}_2)$$

$$= \underset{\sim}{x}'(\Sigma_1^{-1} + \Sigma_2^{-1})\underset{\sim}{x} - 2\underset{\sim}{x}'(\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2) + \underset{\sim}{\mu}_1'\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2'\Sigma_2^{-1}\underset{\sim}{\mu}_2$$

$$= \left[\underset{\sim}{x} - V(\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2)\right]' V^{-1} \left[\underset{\sim}{x} - V(\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2)\right]$$

$$- (\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2)' V (\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \Sigma_2^{-1}\underset{\sim}{\mu}_2) + \underset{\sim}{\mu}_1'\Sigma_1^{-1}\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2'\Sigma_2^{-1}\underset{\sim}{\mu}_2$$

where $V = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1} = \Sigma_{12}$

This expression is the same as -2 times the exponent of the $N_{\underset{\sim}{x}}(\mu_{12},\Sigma_{12})$ except for a constant term . The three components of this constant term can also be

re–written as

$$\text{(a)} \quad \mu_1'\left[\Sigma_1^{-1} - \Sigma_1^{-1}V\Sigma_1^{-1}\right]\mu_1 = \mu_1'(\Sigma_1 + \Sigma_2)^{-1}\mu_1$$

$$\text{(b)} \quad \mu_2'\left[\Sigma_2^{-1} - \Sigma_2^{-1}V\Sigma_2^{-1}\right]\mu_2 = \mu_2'(\Sigma_1 + \Sigma_2)^{-1}\mu_2$$

$$\text{(c)} \quad -2\mu_1'\Sigma_1^{-1}(\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\Sigma_2^{-1}\mu_2 = -2\mu_1'(\Sigma_1 + \Sigma_2)^{-1}\mu_2$$

The equations in part (a) and (b) are due to the matrix inversion lemma and that in (c) holds because

$$(\Sigma_1 + \Sigma_2)^{-1} = \Sigma_1^{-1}(\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\Sigma_2^{-1}$$

$$\Rightarrow \quad \Sigma_1 + \Sigma_2 = \Sigma_2(\Sigma_1^{-1} + \Sigma_2^{-1})\Sigma_1$$
$$= \Sigma_1 + \Sigma_2 \qquad \text{(Q.E.D.)}$$

The constant term works out to be the exponent of $N_{\mu_1}(\mu_2 , \Sigma_1+\Sigma_2)$ multiplied by a factor of –2 . We have , thus , shown that both the L.H.S. and the R.H.S. have the same exponent .

Hence , the properties of the normal density function hold for the multivariate case . As a result , the recursive relations in the Gaussian sum filter extends readily to the multi–state process . The method of collapsing described earlier can be applied here to curb the proliferation of normal terms in the sum of each density . It follows that the AGSF encompasses in a straight forward manner the multivariate approach .

3.3 Parameter Estimation by AGSF

This chapter has so far been concerned with robust filtering , that is , sequential estimation of the state when the parameters of the model are assumed to be known . We shall now turn our attention to the estimation of the variances of the disturbance terms ϵ_t and η_t . This involves the hyperparameters σ_ϵ^2 and σ_η^2 as well as the level of contamination which is determined by g , b , γ and β . Suppose we are analysing a series that has both additive and innovation outliers . Maximum likelihood estimation of the hyperparameters under the assumption of normality will result in over–estimation of σ_ϵ^2 and σ_η^2 . There is a positive bias because the Kalman filter is trying to allow for the outliers in the system . In other words , it will attempt to estimate the variance of the contaminated normal distribution and does not separate the contamination from the true values of the hyperparameters . ML estimation using the AGSF is , therefore , recommended .

3.3.1 The likelihood function

The likelihood of the observations can be expressed as a product of conditional densities $p(y_t/Y_{t-1})$, see equation (2.2.7) . Each of these conditional densities is itself a convolution of the state prediction density and the observation noise density , see equation (2.1.5) . To achieve an accurate representation of the likelihood , the uncollapsed prediction at each stage is used in the convolution . Hence , the four Gaussian terms in (3.2.2) are convoluted with the contaminated normal density in (3.1.2a) . It follows that each conditional density is represented by a normal mixture with eight components . Thus , at time t , we have

$$p(y_t/Y_{t-1}) = \sum_{i=1}^{8} w_i N(\mu_i, \sigma_i^2) \qquad (3.3.1)$$

where

(i) the weights w_i for i=1,2,...,8 are

$(1-\delta_t)(1-\beta)(1-\gamma)$,

$(1-\delta_t)(1-\beta)\,\gamma$,

$(1-\delta_t)\,\beta\,(1-\gamma)$,

$(1-\delta_t)\,\beta\,\gamma$,

$\delta_t\,(1-\beta)(1-\gamma)$,

$\delta_t\,(1-\beta)\,\gamma$,

$\delta_t\,\beta\,(1-\gamma)$

and $\delta_t\,\beta\,\gamma$ respectively .

(ii) the first four components have a common mean $ha_{t/t,1}$ while the last four terms have common mean $ha_{t/t,2}$. Hence , we have

$$\mu_i = \begin{cases} ha_{t/t,1} & \text{for i=1,2,3,4} \\ ha_{t/t,2} & \text{for i=5,6,7,8} \end{cases}$$

and (iii) the variances σ_i^2 for i=1,2,...,8 are

$$h^2 P_{t/t,1} + \sigma_\eta^2 + \sigma_\epsilon^2 \ ,$$

$$h^2 P_{t/t,1} + \sigma_\eta^2 + g\sigma_\epsilon^2 \ ,$$

$$h^2 P_{t/t,1} + b\sigma_\eta^2 + \sigma_\epsilon^2 \ ,$$

$$h^2P_{t/t,1} + b\sigma_\eta^2 + g\sigma_\epsilon^2 ,$$

$$h^2P_{t/t,2} + \sigma_\eta^2 + \sigma_\epsilon^2 ,$$

$$h^2P_{t/t,2} + \sigma_\eta^2 + g\sigma_\epsilon^2 ,$$

$$h^2P_{t/t,2} + b\sigma_\eta^2 + \sigma_\epsilon^2$$

and $h^2P_{t/t,2} + b\sigma_\eta^2 + g\sigma_\epsilon^2$ respectively.

The computed values of ln $p(y_t/Y_{t-1})$ at each stage are added up to give the approximation to the log–likelihood of all the observations

First , let us assume the level of contamination in the model to be known . Then , we can evaluate the MLE of the hyperparameters by maximising over the log–likelihood function . Since the level of contamination is seldom known in practice , we shall investigate its effect on our hyperparameter estimates in the next section and see how these factors of contamination can be estimated in section 3.3.3 .

3.3.2 Contour Maps for the likelihood function

In order to concentrate on the estimation of the hyperparameters , we set $z = h = 1$. That is , we reduce the linear system in (3.1.1) to a local level model . Both disturbances are generated from the contaminated normal distributions with $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$, $\gamma = \beta = 5\%$ and $g = b = 10$. A series of length a hundred is generated from this model . We then run the AGSF with correctly specified level of contamination on

it to obtain a grid of likelihood values in terms of σ_ϵ^2 and σ_η^2 . Figure 3.5a shows a contour map of the log–likelihood curve as a function of the hyperparameters . Note that the function is smooth and that it has global maximum at (0.80,1.31) . Such a well–behaved likelihood implies that there is no starting value problem for the optimisation procedure . This means during the optimisation procedure, the estimates of the hyperparameters will converge to the optimal point even when started at different places .

Other grids for the likelihood function have been obtained for the same series but with misspecifications of the model when running the filter . Figures 3.5b and 3.5c are the contour maps of the log–likelihood function when the level of contamination are misspecified in the filter as (a) $\gamma = \beta = 5\%$, g = b = 1 and (b) $\gamma = \beta = 5\%$, g = b = 100 respectively . That is , the factors g and b are set roughly ten times smaller in case (a) and ten times bigger in case (b) than their actual value . In comparison with figure 3.4a , the log–likelihood curve shifts towards the right in case (a) and it shifts towards the left in case (b) . This usually results in an over–estimation of the hyperparameters in case (a) and an under–estimation of the hyperparameters in case (b) . The optimal points for the two cases are (1.18,1.36) and (0.63,1.46) respectively .

Consider the overall variance of the disturbance terms , that is , the hyperparameters together with the variance of the contaminating components. We can express these terms as follows

$$V(\epsilon_t) = (1-\gamma)\sigma_\epsilon^2 + \gamma g\sigma_\epsilon^2$$

$$= (1 - \gamma + \gamma g)\sigma_\epsilon^2 \qquad (3.3.2a)$$

and $V(\eta_t) = (1-\beta+\beta b)\sigma_\eta^2$ (3.3.2b)

The underlying model from which the data is generated has $V(\epsilon_t) = 1.45\sigma_\epsilon^2$ and $V(\eta_t) = 1.45\sigma_\eta^2$. Assuming the specification of case (a) , the corresponding values becomes $V(\epsilon_t) = \sigma_\epsilon^2$ and $V(\eta_t) = \sigma_\eta^2$ while for (b) , $V(\epsilon_t) = 5.95\sigma_\epsilon^2$ and $V(\eta_t) = 5.95\sigma_\eta^2$. The optimisation procedure is actually estimating the overall variance of the noise terms . Hence , there is likely to be an over–estimation of σ_ϵ^2 and σ_η^2 in case (a) and an under–estimation in case (b) . It is , therefore , necessary for us to identify the true level of contamination in order to estimate the central components only .

3.3.3 Estimation of the factors of contamination

Suppose the percentages of contamination , γ and β , are known . We concentrate on the estimation of the factors of contamination , g and b . The likelihood as a function of these parameters is rather flat . This is because the factors of contamination only affect the series a small proportion of time . Besides , the hyperparameters , σ_ϵ^2 and σ_η^2 , are increased only by a factor of $(1-\gamma+\gamma g)$ and $(1-\beta+\beta b)$ instead of g and b when there is contamination , see equation (3.3.2) . Hence , the inclusion of these factors of contamination into the numerical optimisation procedure for maximum likelihood estimation is not a viable proposition . A more suitable method for identifying the levels of g and b is to use a broad grid search . Suppose that initial estimates of the factors of contamination , denoted by g_0 and b_0 , are available . A scale of ten can be used to form the grid , that is , we have $\frac{1}{10} g_0$, g_0 and $10\, g_0$ spanning the horizontal axis and $\frac{1}{10} b_0$, b_0 and $10\, b_0$ spanning the vertical axis . The ML estimation procedure for the hyperparameters is repeated over

this broad grid of values for g and b . By comparing the ML value for each combination , the best model is chosen as the one with the biggest likelihood value . We note that the grid search for the factors of contamination is similar to the model identification method employed by Kitagawa (1987) which uses the Akaike information criterion .

We could try a slightly different version of the above AGSF . Suppose that no collapsing was done at the prediction stage , i.e. , all four terms of the prediction density are used in the updating process . The state updating density will end up with eight terms which are combined to give a two term mixture . Like the previous filter , the normalising constant $p(y_t/Y_{t-1})$ which contributes towards the likelihood value is also made up of eight terms . However , these terms are different from those in the previous filter . In the latter , a distinction is made between outliers and non–outliers , leading to better estimates of the state . It seems a bit surprising that the model selection process using the AIC does not work with this new version of AGSF . Hence , we conclude that the identification of aberrant points at the prediction stage helps to make the likelihood function more sensitive towards the specification of the contaminating component in the model .

Initial estimates for the factors of contamination can be evaluated with the procedures used to obtain starting values for the hyperparameters . The latter are usually derived from the sample variance and covariance of a differenced series . Taking first differences of the observations , we have from a local level model ,

$$\begin{aligned} \Delta y_t &= y_t - y_{t-1} \\ &= \eta_t + \epsilon_t - \epsilon_{t-1} \end{aligned} \qquad (3.3.3)$$

Thus , the variance and covariance of this differenced series can be written as

$$\text{Var}(\Delta\, y_t) = \sigma_\eta^2 + 2\,\sigma_\epsilon^2 \tag{3.3.4a}$$

and $$\text{Cov}(\Delta\, y_t, \Delta\, y_{t+1}) = -\,\sigma_\epsilon^2 \tag{3.3.4b}$$

Such functional relations enable us to obtain rough estimates of the hyperparameters based on the sample estimates of $V(\Delta\, y_t)$ and $\text{Cov}(\Delta\, y_t, \Delta\, y_{t+1})$. Due to presence of outliers , it is necessary for the latter to be robust . Some trimming of proportion α can be introduced . Hence , we have

$$\hat{V}(\Delta\, y_t) = \frac{\Sigma\,(\Delta y_t - \Delta\tilde{y})^2}{(1-\alpha)(T-1)} \tag{3.3.5a}$$

$$\hat{\text{Cov}}(\Delta\, y_t, \Delta\, y_{t+1}) = \frac{\Sigma\,(\Delta y_t - \Delta\tilde{y})(\Delta y_{t+1} - \Delta\tilde{y})}{(1-\alpha)(T-2)} \tag{3.3.5b}$$

where $\Delta\tilde{y}$ is the median of the differenced observations and the summations exclude $\frac{\alpha}{2}$ proportion of the largest and $\frac{\alpha}{2}$ proportion of the smallest summants . On the other hand , if we use all the differenced observations and compute $\hat{V}(\Delta y_t)$ and $\hat{\text{Cov}}(\Delta y_t, \Delta y_{t+1})$ as the sample variance and covariance , we end up with rough estimates of the overall variances . Using the relations in equation (3.3.2) , we can then work out approximate values of g and b from these initial estimates of $V(\epsilon_t)$, $V(\eta_t)$, σ_ϵ^2 and σ_η^2 .

3.3.4. Simulation results

In order to study the performance of the MLEs from the AGSF and compare them with those from the Kalman filter , we carry out some simulations . Series of length one hundred are generated from the local level model . When generating the data , the first thirty observations are discarded to remove the effect of the initial state on the series . An optimisation routine C05NBF from the NAG library is applied to carry out the ML estimation for both the AGSF and the Kalman filter . One hundred replications are used in each set of simulations . The summary statistics of the estimates , namely the bias and mean square errors (MSE) , taken over these one hundred replications are then evaluated . We shall investigate the sampling properties of the two types of estimators for the following set of parameter values :

(a) $\sigma_\epsilon^2 = 2$, $\sigma_\eta^2 = 1$;
(b) $\sigma_\epsilon^2 = 2$, $\sigma_\eta^2 = 2$; and
(c) $\sigma_\epsilon^2 = 1$, $\sigma_\eta^2 = 1.5$.

In order to facilitate the comparison amongst the three cases , we perform a logarithmic transformation on the hyperparameters . Hence , we estimate the logarithms of the scales , i.e. ln σ_ϵ and ln σ_η , in the three sets of simulations . If there is at least one breakdown in any set of simulations , another set of summary statistics which are more resistant is computed . The bias , denoted by bias* , is now evaluated using the median while the mean square error , denoted by MSE* , is computed using the pseudovariance , see Hoaglin et al (1983) . The latter is the square of the interquartile range scaled by 1.349 which makes it comparable with the standard deviation of the Gaussian distribution . We say that there is a breakdown

in the estimation when the estimate of a parameter in at least one replication is way off from most other estimates within the same simulation . The number of cases which are excluded , M , is also indicated when presenting this alternative set of summary statistics .

When comparing the summary statistics of two different estimation procedure, we have to ensure that any difference found in the bias and mean square error are not due to sampling variability . Suppose that the sample estimate of the mean μ and variance σ^2 of a distribution are $\bar{x}$ and $\hat{\sigma}^2$ respectively . Then , assuming normality , we have

$$\text{var}(\bar{x}) \approx \frac{\hat{\sigma}^2}{n} \tag{3.3.6a}$$

and

$$\text{var}(\hat{\sigma}^2) \approx 2\frac{\hat{\sigma}^4}{n-1} \tag{4.5.6b}$$

Denoting the bias and mean square error of the distribution by β and ϕ respectively , it follows that

$$\text{var}(\beta) \approx \frac{\hat{\sigma}^2}{n} \tag{4.3.7a}$$

and

$$\text{var}(\phi) \approx 2\hat{\sigma}^4 \left[\frac{1}{n^2} + \frac{1}{n-1}\right] \tag{4.3.7b}$$

To check if the difference in the summary statistic is significant , we need to consider the variance of this difference . Independence of the simulations is assumed so that this variance is given by the sum of the variance from each sampling distribution . Once , we obtain the following statistics

$$z = \frac{\text{difference in summary statistics}}{\text{standard deviation of difference}}$$

, we can compare it with the z_p value from the normal table . We conclude that the difference is significant at the pth level when $z > z_p$. In our analysis of the simulation results , we shall fix p at the 5% level , unless otherwise stated .

The simulation results are categorised according to the types of contamination present in the system . We shall consider four different cases , namely , clean data ; additive outliers only ; innovation outliers only ; and both additive and innovation outliers . These cases can be modelled by the distribution of the disturbance terms , see equation (3.1.2) . For simplicity , let us fix the percentage of contamination , both γ and β , at 5% and assume that to be known . Then , the above four types of contamination are produced by setting :

Case 1 $g = b = 1$;

Case 2 $g = 20$, $b = 1$;

Case 3 $g = 1$, $b = 10$; and

Case 4 $g = 20$, $b = 10$.

For each case , we shall first present the summary statistics from the AGSF and then from the Kalman filter .

Case 1 Clean data

TABLE 3.1 SUMMARY STATISTICS FROM AGSF (clean data)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	–0.0213	0.0175
(b)	0.3466	–0.0296	0.0285
(c)	0.0	0.0055	0.0041
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.0647	0.0672
(b)	0.3466	–0.0661	0.0552
(c)	0.2027	–0.0906	0.0459

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.6 to 3.8 .

TABLE 3.2 SUMMARY STATISTICS FROM KALMAN FILTER (clean data)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	–0.0152	0.0129
(b)	0.3466	–0.0258	0.0221
(c)	0.0	–0.0053	0.0332
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.0165	0.0496
(b)	0.3466	–0.0121	0.0308
(c)	0.2027	–0.0504	0.0371

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.9 to 3.11 .

We observe , in the absence of contamination , that the MLE of ln σ_ϵ from the two filters are not significantly different in terms of their summary statistics . However , we can see more clearly the edge which the Kalman filter estimator has over that produced by the AGSF for the parameter ln σ_η . In case (b) , where $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$, both the bias and mean square error for the MLE of ln σ_η from the Kalman filter are significantly smaller than those from the AGSF . We expect slightly better behaviour for the estimators from the Kalman filter when there are no outliers in the data because unlike the AGSF , it does not involve the estimation of the levels of contamination .

Case 2 Additive outliers only

TABLE 3.3 SUMMARY STATISTICS FROM AGSF (AO only)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	0.0329	0.0168
(b)	0.3466	0.0417	0.0292
(c)	0.0	0.0393	0.0363
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.1617	0.1263
(b)	0.3466	–0.1071	0.0661
(c)	0.2027	0.0811	0.0476

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.12 to 3.14 .

TABLE 3.4 SUMMARY STATISTICS FROM KALMAN FILTER (AO only)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	0.3001	0.1358
(b)	0.3466	0.2967	0.1389
(c)	0.0	0.3005	0.1449
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.1071	0.0740
(b)	0.3466	–0.0754	0.0529
(c)	0.2027	–0.0695	0.0457

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.15 to 3.17 .

When additive outliers are present in the series , we find that the AGSF can estimate ln σ_ϵ better than the Kalman filter . This is evident from Tables 3.3 and 3.4 where we see the summary statistics for the MLE from the Kalman filter are significantly bigger than those from the AGSF in all three cases . Since outliers are absent from the transition equation , the parameter ln σ_η is still quite well estimated by the Kalman filter . The summary statistics for lnσ_η from the two filters are not very different except in case (a) where $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$, where we see a significantly smaller MSE for the estimator from the Kalman filter . Hence , we conclude that the AGSF is more robust towards additive outliers compared with the Kalman filter .

Case 3 Innovation outliers only

TABLE 3.5 SUMMARY STATISTICS FROM AGSF (IO only)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	–0.0181	0.0178
(b)	0.3466	0.0122	0.0486
(c)	0.0	0.0348	0.0494
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.0183	0.1028
(b)	0.3466	0.0190	0.0597
(c)	0.2027	–0.0052	0.0532

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.18 to 3.20 .

TABLE 3.6 SUMMARY STATISTICS FROM KALMAN FILTER (IO only)

Estimator of		Bias	MSE	Bias*	MSE*	
(i)	$\log \sigma_\epsilon$					
(a)	0.3466	–0.0116	0.0161			
(b)	0.3466	–0.1163	0.6569	0.0104	0.0240	2
(c)	0.0	–0.6620	6.2078	0.0180	0.0380	9
(ii)	$\log \sigma_\eta$					
(a)	0.0	0.1454	0.0660			
(b)	0.3466	0.1237	0.0804			
(c)	0.2027	0.1374	0.0817			

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.21 to 3.23 .

The presence of innovation outliers in the system not only affects the estimation of ln σ_η by the Kalman filter but also causes its estimation of ln σ_ϵ to sometimes breakdown . Such breakdown in the estimation of ln σ_ϵ does not occur with the AGSF . In case (a) where $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$, the MLE of ln σ_ϵ from the Kalman filter does not break down and there is no significant difference in its summary statistics compared with corresponding values from the AGSF . With the exception of the MSE for case (a) , we see that for the parameter ln σ_η , the bias as well as the mean square error obtained from the Kalman filter are significantly bigger than corresponding values produced by the AGSF in all three cases . This leads to the conclusion that unlike the Kalman filter , the AGSF is rather insensitive towards innovation outliers .

Case 4 Additive and Innovation outliers

TABLE 3.7 SUMMARY STATISTICS FROM AGSF (AO and IO)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	0.0329	0.0197
(b)	0.3466	0.0268	0.0386
(c)	0.0	0.0348	0.0494
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.0433	0.1159
(b)	0.3466	0.0059	0.0651
(c)	0.2027	0.0052	0.0532

The sampling distributions of the two MLEs from the two filters for the three different sets of parameter values are given in figures 3.24 to 3.26 .

TABLE 3.8 SUMMARY STATISTICS FROM KALMAN FILTER (AO and IO)

Estimator of		Bias	MSE	Bias*	MSE*	
(i)	$\log \sigma_\epsilon$					
(a)	0.3466	0.2991	0.1371			
(b)	0.3466	0.2909	0.1426			
(c)	0.0	0.2357	0.4765	0.2985	0.6014	1
(ii)	$\log \sigma_\eta$					
(a)	0.0	0.0797	0.0903			
(b)	0.3466	0.1039	0.0848			
(c)	0.2027	0.1101	0.0807			

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.27 to 3.29 .

Here , both the measurement and transition equation of the system are contaminated by aberrant data . The AGSF shows a definite edge over the Kalman filter in the estimation of ln σ_ϵ . Both the bias and mean square error from the Kalman filter are significantly bigger than corresponding values from the AGSF . In case (c) , where $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$, there is a breakdown in the estimation of ln σ_ϵ by the Kalman filter . The improvement made by using the AGSF is less obvious in the estimation of ln σ_η . This is partly because a smaller factor of contamination is used , i.e. b < g . In case (a) , there is no significant difference in the summary statistics of the two filters . However , the summary statistics from the Kalman filter for the other two cases are significantly bigger than those produced by the AGSF .

Thus , we say that the AGSF can cope with the presence of both additive and innovation outliers better in terms of the estimation of the hyperparameters than the Kalman filter .

Consider the ratio of the system noise variance σ_η^2 to the observation noise variance σ_ϵ^2 and denote this quantity by q . In the above simulations , we have considered different sets of parameter values with q assuming the values 0.5 , 1 and 1.5 . When the value of q is big , it seems as if the MLE procedure for estimating σ_ϵ tends to break down . This is because scale is a positive parameter so that negative estimates are given the value zero . This results in a lump at the zero level for the sampling distribution of the MLE . In the logarithm scale , this corresponds to large negative estimates for ln σ_ϵ . Let us consider all one hundred replications in the simulation for case (c) when the data is clean . Figures 3.30 and 3.31 show the sampling distributions for the MLEs of the ln σ_ϵ from the Kalman filter and the AGSF respectively . Two cases of breakdown at replications 41 and 57 can be identified for the Kalman filter estimator . Corresponding estimates from the AGSF at these two replications reveal that with the AGSF breakdowns occur less frequently and less drastically . In order not to confound this effect with the distortion caused by outliers , we excluded the estimates from these replications for both filters in all the above simulations for case (c) .

3.3.5 Some Examples

Recall the two examples given in section 3.2.3 . Suppose we now run the AGSF and the Kalman filter on the generated series without assuming knowledge of the parameter values . That is , we have to perform ML estimation of the

hyperparameters before tracking the state . For the AGSF , a broad grid search of the factors of contamination is included though the percentage of contamination is set correctly at the 5% level .

Let us first consider the series with only additive outliers , i.e. g=100 , b=1 . The scale , namely , σ_ϵ and σ_η are estimated as 1.05 and 0.91 respectively by the AGSF while the Kalman filter's estimates are 1.75 and 1.00 respectively . A plot of the generated state α_t and its estimates $a_{t/t-1}$ from both filters are found in figure 3.32 . We observe that the AGSF , being more robust towards outliers , can track the state better than the Kalman filter . The other series has noisy disturbances at both the measurement and transition equations , i.e. g=100 and b=10 . Estimates of scale σ_ϵ and σ_η from the AGSF are 1.04 and 0.93 respectively . Corresponding values produced by the Kalman filter are 1.74 and 1.01 respectively . Figure 3.33 consist of a plot of the generated state and its estimates from the two filter . We conclude that the AGSF can still track the state better than the Kalman filter even in the presence of noisy state disturbance term . We observe that with the estimation of hyperparameters , the Kalman filter is not as sensitive to additive outliers , compare figures 3.32 and 3.3 . This is because its over–estimation of σ_ϵ will lead to a smaller estimate of the ratio q which results in slower reaction of the filter to the observations .

Next , let us consider a real example . We have a set of monthly data on UK retail price index from 1968 to 1984 . We first difference the series and then fit a local level model to it . Any step changes in the original series will become additive outliers after differencing . Running the AGSF on this series produces estimates of scale σ_ϵ and σ_η as 3.57 and 0.77 respectively . Corresponding estimates from the Kalman filter are 7.61 and 0.99 . Figure 3.34 gives a plot of the series y_t and the

one–step ahead predictions $a_{t/t-1}$ from the two filters . We observe that the Kalman filter is more sensitive towards the aberrant observations , see the 76th and 132nd observation which correspond to prices in 1974 and 1979 . On the other hand , the state predictions from the AGSF are more robust towards the outliers .

3.4 Non–normal contamination

So far , we have been generating outliers by allowing the distribution of the disturbance terms take on contaminated normal distributions , see equation (3.1.2) . However , other forms of long–tailed distribution may also be reasonable approximations to reality . We shall let the central component of the disturbance terms remain as normal variates but use a t–distribution with two degrees of freedom for the contaminating components . That is , a series with both additive and innovation outliers is generated by having

$$p(\epsilon_t) = (1-\gamma)\, N(0,\sigma_\epsilon^2) + \gamma\, \omega_{1t} \qquad (4.4.1a)$$

$$p(\eta_t) = (1-\beta)\, N(0,\sigma_\eta^2) + \beta\, \omega_{2t} \qquad (4.4.1b)$$

where ω_{1t} and $\omega_{2t} \sim t_2$ distribution

To investigate how the two filters , namely , the AGSF and the Kalman filter cope with outliers of this nature , we repeat the simulations for case 4 where contamination occurs at both equations of the system . The summary statistics for

the three sets of parameter values are given in Tables 3.9 and 3.10 below .

TABLE 3.9 SUMMARY STATISTICS FROM KALMAN FILTER (t_2 distribution)

Estimator of		Bias	MSE	Bias*	MSE*	
(i)	$\log \sigma_\epsilon$					
(a)	0.3466	–0.0432	0.3110	–0.0002	0.0146	1
(b)	0.3466	–0.2406	2.6564	–0.0026	0.0190	3
(c)	0.0	–0.4210	3.7872	0.0150	0.0382	7
(ii)	$\log \sigma_\eta$					
(a)	0.0	0.0541	0.0848			
(b)	0.3466	0.0248	0.0587			
(c)	0.2027	–0.0628	0.0667			

The sampling distributions of these two estimators for the three different sets of parameter values are given in figures 3.35 to 3.37 .

TABLE 3.10 SUMMARY STATISTICS FROM AGSF (t_2 distribution)

Estimator of		Bias	MSE
(i)	$\log \sigma_\epsilon$		
(a)	0.3466	–0.0121	0.0142
(b)	0.3466	–0.0216	0.0272
(c)	0.0	–0.0441	0.0639
(ii)	$\log \sigma_\eta$		
(a)	0.0	–0.0855	0.1066
(b)	0.3466	–0.0588	0.0493
(c)	0.2027	–0.0341	0.0367

The sampling distributions of these two MLEs for the three different sets of parameter values are given in figures 3.38 to 3.40 .

The presence of non–normal contamination causes the Kalman filter to breakdown in its estimation of ln σ_ϵ for all three cases . Breakdown occurs for the estimation of ln σ_η by the Kalman filter for only case (c) . Estimation of neither hyperparameter broke down for the AGSF . The summary statistics for the estimator of ln σ_η from the AGSF are not significantly different from those produced by the Kalman filter in cases (a) and (b) . Hence , we conclude that the AGSF is more robust when compared with the Kalman filter towards non–normal contamination . In fact , the performance of the MLE from the AGSF would improve as we increase the number of terms in the Gaussian sums . This is because we can achieve a more accurate representation of a non–normal component with more terms in the normal mixture . But this will lead to an increase in the computation time .

There is , thus , a trade off between good representation and computation time involved . In most economic time series data , it is enough to model the outliers using a normal contaminating component since wild outliers can usually be identified and removed during a preliminary examination of the plot of the series .

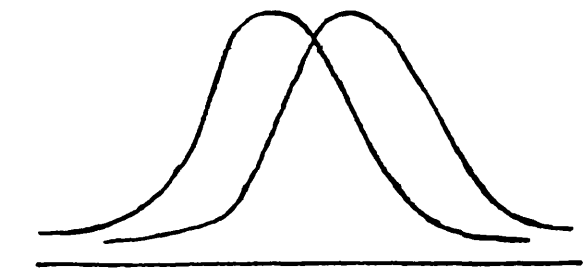


Figure 3.1a Mixture of two close components

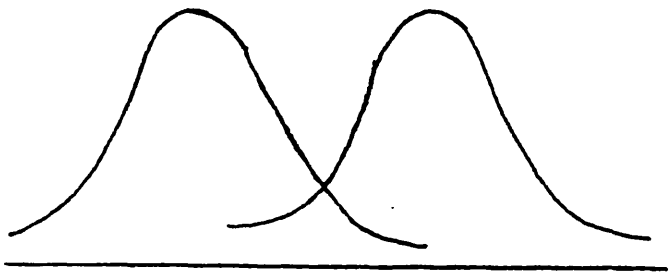
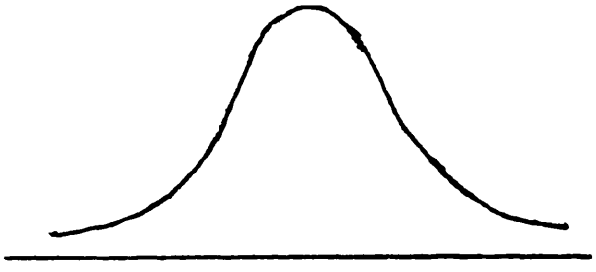
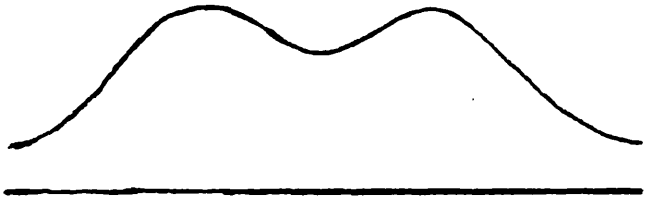


Figure 3.1b Mixture of two well–separated components





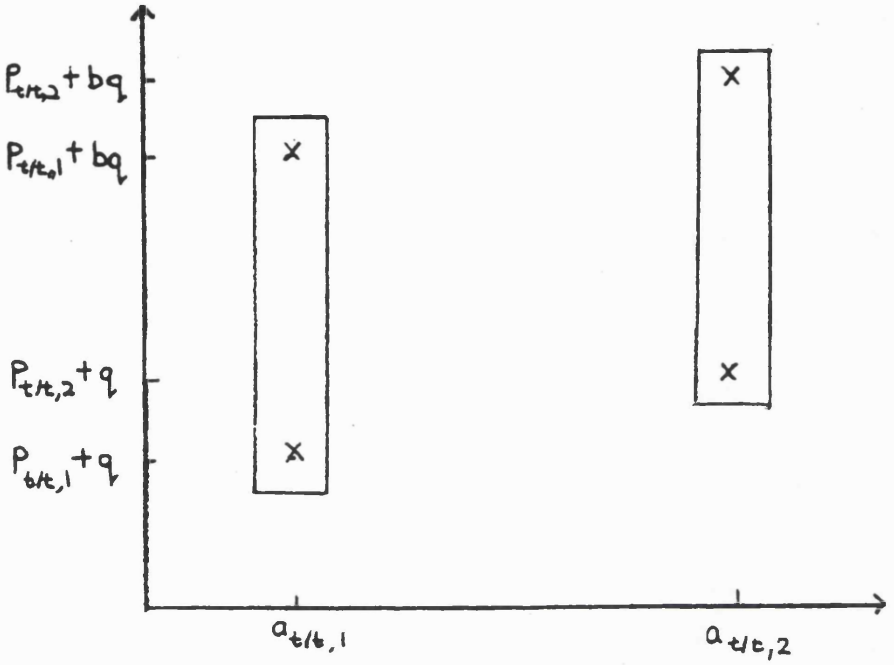


Figure 3.2a Asymmetric collapsed density

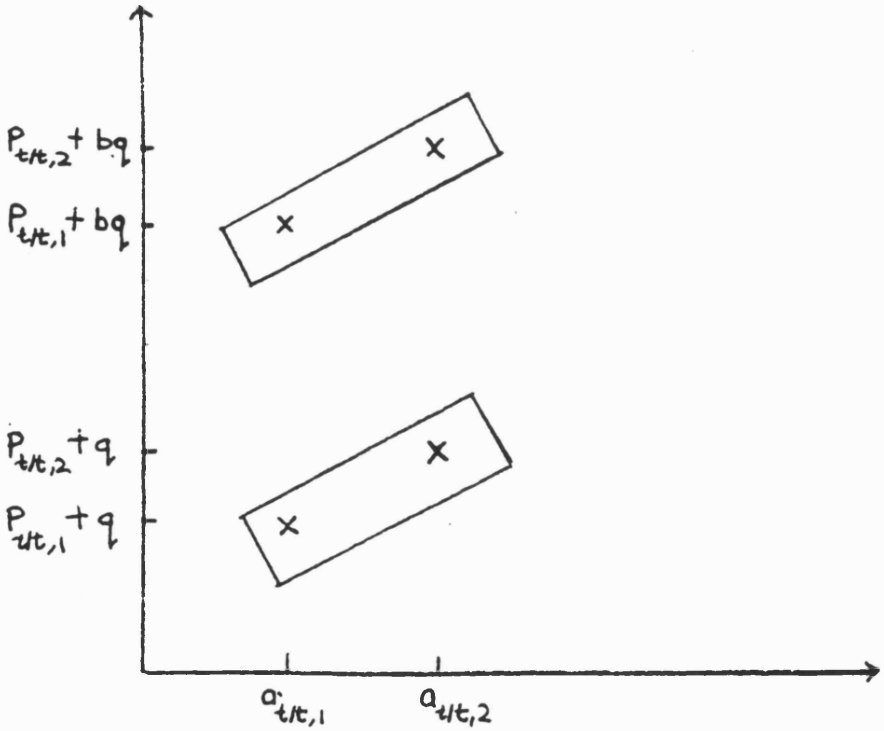


Figure 3.2b Symmetric collapsed density

Tracking of state (———) by AGSF (— —) and Kalman filter (– – –) without parameter estimation

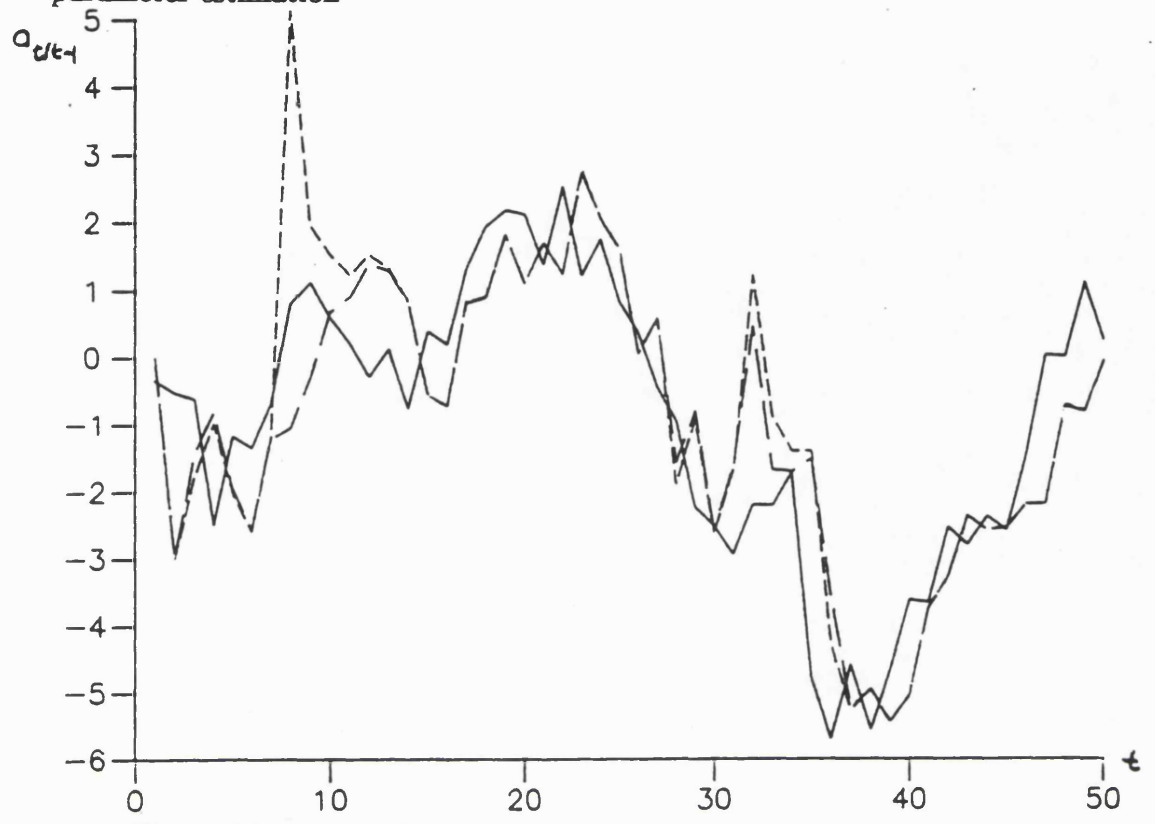


Figure 3.3 $g = 100$, $b = 1$

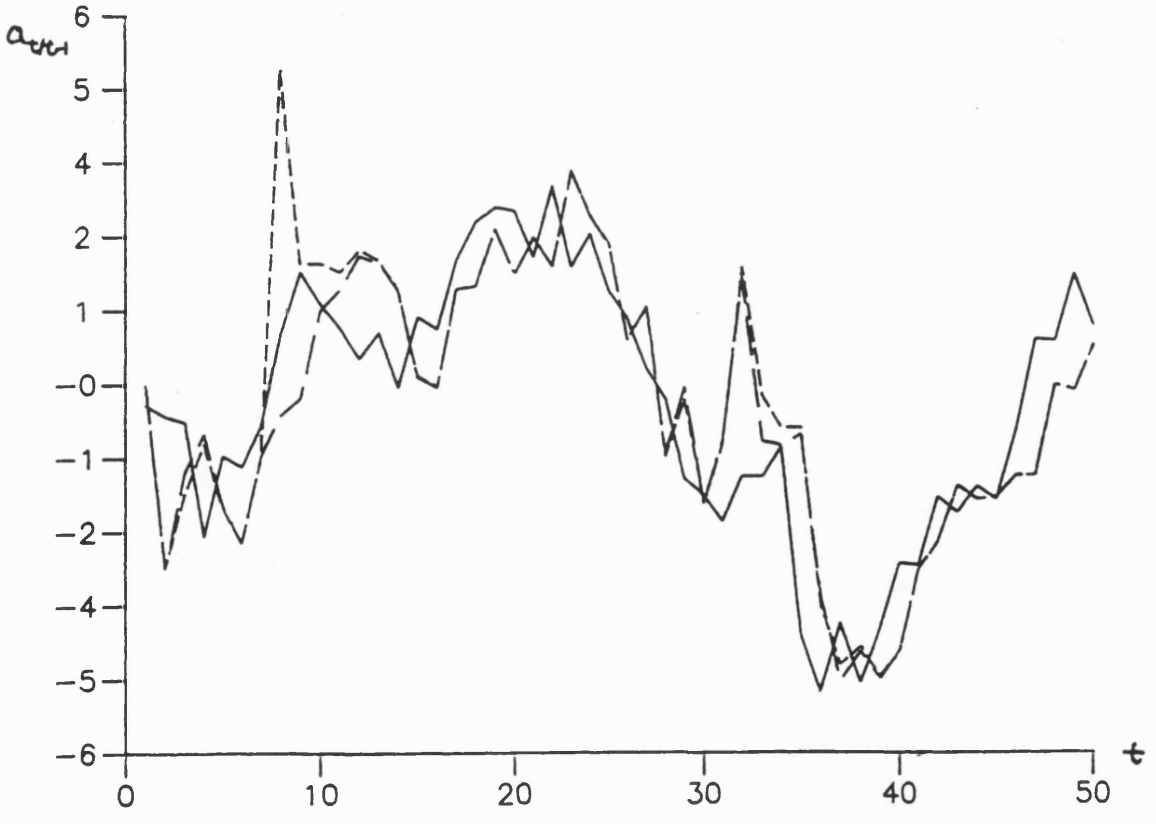


Figure 3.4 $g = 100$, $b = 10$

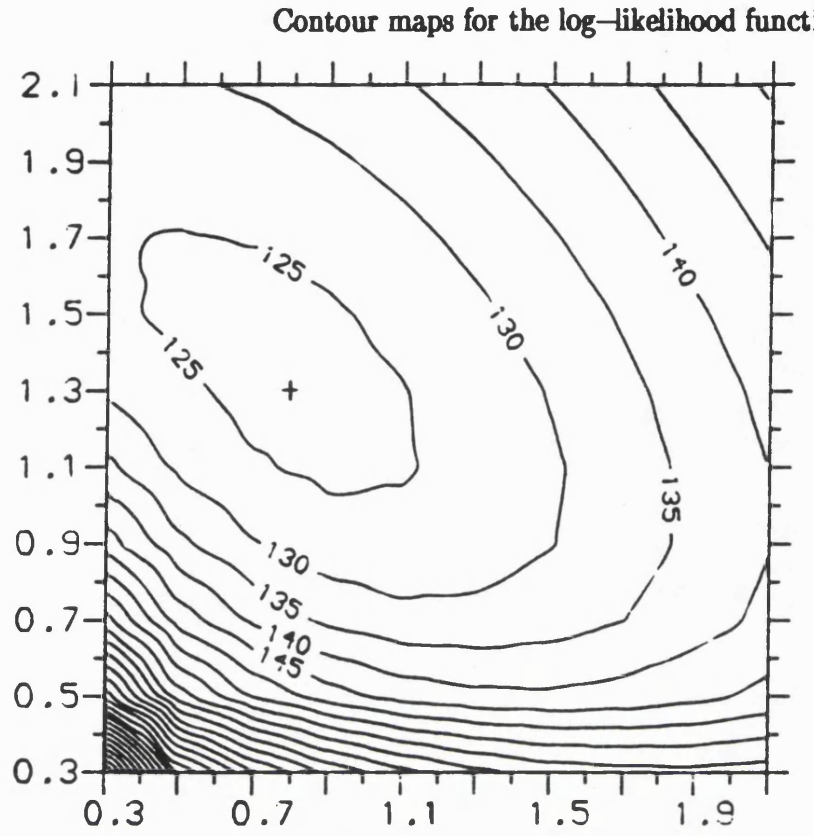


Figure 3. 5a Correct specification of GK and BK

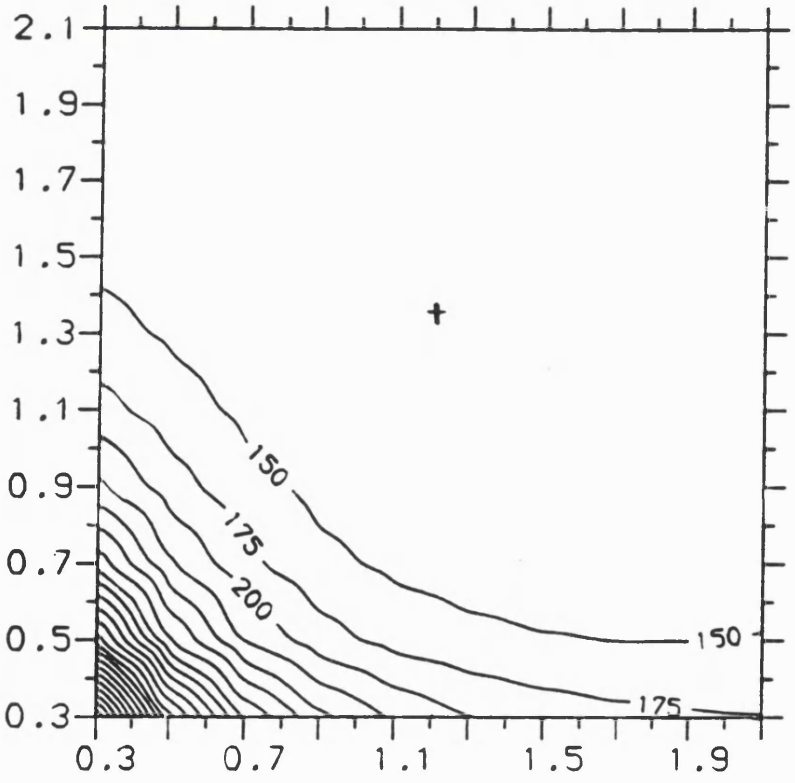


Figure 3. 5b Under–specification of GK and BK

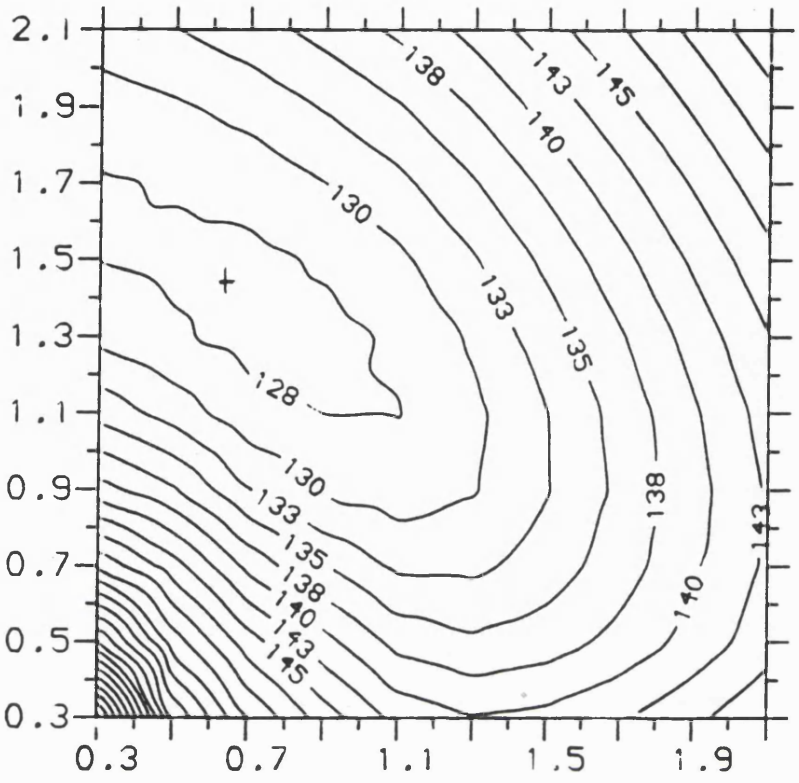


Figure 3. 5c Over–specification of GK and BK

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln σ_η from AGSF for clean data

(a) Midpoint Count

−0.1	2	**
0.0	1	*
0.1	6	******
0.2	16	**********
0.3	29	**********
0.4	31	**********
0.5	13	**********
0.6	2	**

(b) Midpoint Count

−1.4	1	*
−1.2	0	
−1.0	0	
−0.8	2	**
−0.6	2	**
−0.4	8	********
−0.2	25	**********
0.0	39	**********
0.2	17	**********
0.4	6	*******

Figure 3.6 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

−0.3	1	*
−0.2	1	*
−0.1	0	
0.0	3	***
0.1	9	*********
0.2	16	**********
0.3	25	**********
0.4	22	**********
0.5	16	**********
0.6	7	******

(b) Midpoint Count

–0.7	1	*
–0.6	0	
–0.5	0	
–0.4	2	**
–0.3	0	
–0.2	2	**
–0.1	2	**
0.0	2	**
0.1	11	**********
0.2	20	********************
0.3	18	******************
0.4	23	************************
0.5	10	**********
0.6	9	*********

Figure 3.7 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.4	3	***
–0.3	6	******
–0.2	9	*********
–0.1	12	************
0.0	20	********************
0.1	35	***********************************
0.2	9	*********
0.3	4	****

(b) Midpoint Count

–0.5	1	*
–0.4	2	**
–0.3	0	
–0.2	6	******
–0.1	11	**********
0.0	12	************
0.1	20	********************
0.2	22	**********************
0.3	16	****************
0.4	8	********

Figure 3.8 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln σ_η from Kalman filter for clean data

(a) Midpoint Count

0.1	8	********
0.2	15	***************
0.3	30	******************************
0.4	34	**********************************
0.5	11	***********
0.6	2	**

(b) Midpoint Count

−1.2	1	*
−1.0	0	
−0.8	0	
−0.6	2	**
−0.4	4	****
−0.2	24	************************
0.0	36	************************************
0.2	28	****************************
0.4	5	******

Figure 3.9 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

0.0	5	*****
0.1	9	*********
0.2	14	**************
0.3	28	****************************
0.4	22	**********************
0.5	16	****************
0.6	6	******

(b) Midpoint Count

−0.4	1	*
−0.3	0	
−0.2	1	*
−0.1	1	*
0.0	2	**
0.1	5	*****
0.2	18	******************
0.3	22	**********************
0.4	27	***************************
0.5	15	***************
0.6	7	*******
0.7	1	*

Figure 3.10 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.6	1	*
–0.5	2	**
–0.4	2	**
–0.3	5	*****
–0.2	11	**********
–0.1	9	*********
0.0	24	**********
0.1	29	**********
0.2	10	**********
0.3	5	*****

(b) Midpoint Count

–0.4	2	**
–0.3	1	*
–0.2	3	***
–0.1	11	**********
0.0	6	******
0.1	19	**********
0.2	25	**********
0.3	21	**********
0.4	8	********
0.5	2	**

Figure 3.11 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln σ_η from AGSF for data with additive outliers only

(a) Midpoint Count

0.0	1	*
0.1	5	*****
0.2	8	*******
0.3	21	**********
0.4	39	**********
0.5	17	**********
0.6	9	*********

(b) Midpoint Count

–1.4	2	**
–1.2	1	*
–1.0	0	
–0.8	1	*
–0.6	10	**********
–0.4	9	*********
–0.2	26	**************************
0.0	35	***********************************
0.2	14	**************
0.4	2	**

Figure 3.12 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

–0.1	1	*
0.0	2	**
0.1	8	********
0.2	6	******
0.3	18	******************
0.4	30	******************************
0.5	20	********************
0.6	12	************
0.7	2	**
0.8	1	*

(b) Midpoint Count

–0.4	3	***
–0.3	1	*
–0.2	3	***
–0.1	4	****
0.0	9	*********
0.1	11	***********
0.2	13	*************
0.3	22	**********************
0.4	16	****************
0.5	13	*************
0.6	3	***
0.7	2	**

Figure 3.13 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

−0.4	4	****
−0.3	6	******
−0.2	5	*****
−0.1	9	*********
0.0	25	*************************
0.1	22	**********************
0.2	17	*****************
0.3	8	********
0.4	1	*
0.5	1	*

(b) Midpoint Count

−0.5	3	***
−0.4	0	
−0.3	0	
−0.2	7	*******
−0.1	10	**********
0.0	9	*********
0.1	19	*******************
0.2	21	*********************
0.3	19	*******************
0.4	8	********
0.5	2	**

Figure 3.14 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) $\ln \sigma_\epsilon$ and (b) $\ln \sigma_\eta$ from Kalman filter for data with additive outliers only

(a) Midpoint Count

0.2	2	**
0.3	8	********
0.4	15	***************
0.5	7	*******
0.6	16	****************
0.7	18	******************
0.8	16	****************
0.9	12	************
1.0	4	****
1.1	1	*
1.2	4	****

(b) Midpoint Count

–0.8	2	**
–0.6	5	*****
–0.4	14	****
–0.2	23	**********
0.0	37	**********
0.2	15	**********
0.4	4	****

Figure 3.15 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

0.1	2	**
0.2	3	***
0.3	4	****
0.4	14	**********
0.5	10	**********
0.6	13	**********
0.7	21	**********
0.8	12	**********
0.9	14	**********
1.0	5	*****
1.1	1	*
1.2	1	*

(b) Midpoint Count

–0.4	1	*
–0.3	0	
–0.2	2	**
–0.1	8	********
0.0	5	*****
0.1	10	**********
0.2	13	**********
0.3	21	**********
0.4	21	**********
0.5	12	**********
0.6	4	****
0.7	3	***

Figure 3.16 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.4	1	*
–0.3	1	*
–0.2	2	**
–0.1	3	***
0.0	7	*******
0.1	11	***********
0.2	11	***********
0.3	19	*******************
0.4	17	*****************
0.5	13	*************
0.6	9	*********
0.7	2	**
0.8	2	**

(b) Midpoint Count

–0.6	1	*
–0.5	0	
–0.4	1	*
–0.3	1	*
–0.2	8	********
–0.1	7	*******
0.0	9	*********
0.1	19	*******************
0.2	22	**********************
0.3	18	******************
0.4	9	*********
0.5	3	***

Figure 3.17 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) $\ln \sigma_\epsilon$ and (b) $\ln \sigma_\eta$ from AGSF for data with innovation outliers only

(a) Midpoint Count

0.0	2	**
0.1	7	*******
0.2	22	**********************
0.3	24	************************
0.4	22	**********************
0.5	20	********************
0.6	3	***

(b) Midpoint Count

−1.2	1	*
−1.0	0	
−0.8	4	****
−0.6	4	****
−0.4	4	****
−0.2	16	**********
0.0	31	**********
0.2	29	**********
0.4	10	**********
0.6	1	*

Figure 3.18 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

−0.4	1	*
−0.3	0	
−0.2	0	
−0.1	2	**
0.0	1	*
0.1	5	*****
0.2	12	**********
0.3	23	**********
0.4	29	**********
0.5	19	**********
0.6	7	******

(b) Midpoint Count

−0.3	1	*
−0.2	2	**
−0.1	4	****
0.0	7	*******
0.1	10	**********
0.2	7	*******
0.3	20	**********
0.4	19	**********
0.5	14	**********
0.6	8	********
0.7	4	****
0.8	2	**
0.9	2	**

Figure 3.19 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.6	1	*
–0.5	3	***
–0.4	1	*
–0.3	6	******
–0.2	9	*********
–0.1	6	******
0.0	21	**********************
0.1	20	*********************
0.2	17	*****************
0.3	9	*********
0.4	5	*****

(b) Midpoint Count

–0.5	2	**
–0.4	0	
–0.3	1	*
–0.2	2	**
–0.1	10	**********
0.0	6	******
0.1	15	***************
0.2	19	*******************
0.3	15	***************
0.4	18	******************
0.5	7	*******
0.6	2	**
0.7	1	*

Figure 3.20 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) $\ln \sigma_\epsilon$ and (b) $\ln \sigma_\eta$ from Kalman filter data with innovation outliers only

(a) Midpoint Count

0.0	2	**
0.1	7	*******
0.2	17	*****************
0.3	23	**********************
0.4	33	*********************************
0.5	15	***************
0.6	3	***

(b) Midpoint Count

–0.8	1	*
–0.6	0	
–0.4	2	**
–0.2	8	********
0.0	28	****************************
0.2	38	**************************************
0.4	22	**********************
0.6	1	*

Figure 3.21 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

2 observations below the first class

(a) Midpoint Count

–0.4	1	*
–0.3	0	
–0.2	1	*
–0.1	1	*
0.0	4	****
0.1	6	******
0.2	9	*********
0.3	24	************************
0.4	23	***********************
0.5	17	*****************
0.6	10	**********
0.7	2	**

(b) Midpoint Count

–0.3	1	*
–0.2	0	
–0.1	2	**
0.0	2	**
0.1	8	********
0.2	7	*******
0.3	6	******
0.4	15	***************
0.5	20	********************
0.6	16	****************
0.7	11	***********
0.8	7	*******
0.9	3	***
1.0	1	*
1.1	1	*

Figure 3.22 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

9 observations below the first class

(a) Midpoint Count

–0.6	1	*
–0.5	0	
–0.4	4	****
–0.3	2	**
–0.2	6	******
–0.1	14	**********
0.0	19	**********
0.1	24	**********
0.2	11	**********
0.3	7	*******
0.4	1	*

(b) Midpoint Count

–0.4	1	*
–0.3	0	
–0.2	1	*
–0.1	3	***
0.0	7	*******
0.1	9	*********
0.2	10	**********
0.3	18	**********
0.4	15	**********
0.5	15	**********
0.6	11	**********
0.7	4	****
0.8	2	**
0.9	2	**

Figure 3.23 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln σ_η from AGSF for data with both additive and innovation outliers

(a) Midpoint Count

–0.1	1	*
0.0	1	*
0.1	2	**
0.2	11	**********
0.3	22	**********
0.4	31	**********
0.5	22	**********
0.6	10	**********

(b) Midpoint Count

−1.2	1	*
−1.0	2	**
−0.8	1	*
−0.6	8	********
−0.4	8	********
−0.2	13	*************
0.0	27	***************************
0.2	29	*****************************
0.4	9	*********
0.6	2	**

Figure 3.24 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

−0.2	2	**
−0.1	1	*
0.0	3	***
0.1	8	********
0.2	8	********
0.3	15	***************
0.4	26	**************************
0.5	21	**********************
0.6	10	**********
0.7	6	******

(b) Midpoint Count

−0.4	1	*
−0.3	0	
−0.2	3	***
−0.1	4	****
0.0	5	*****
0.1	9	*********
0.2	7	*******
0.3	15	***************
0.4	19	*******************
0.5	15	***************
0.6	10	**********
0.7	9	*********
0.8	2	**
0.9	1	*

Figure 3.25 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.6	1	*
–0.5	3	***
–0.4	1	*
–0.3	6	******
–0.2	9	*********
–0.1	6	******
0.0	21	**********************
0.1	20	*********************
0.2	17	*****************
0.3	9	*********
0.4	5	*****

(b) Midpoint Count

–0.5	2	**
–0.4	0	
–0.3	1	*
–0.2	2	**
–0.1	10	**********
0.0	6	******
0.1	15	***************
0.2	19	*******************
0.3	15	***************
0.4	18	******************
0.5	7	*******
0.6	2	**
0.7	1	*

Figure 3.26 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) $\ln \sigma_\epsilon$ and (b) $\ln \sigma_\eta$ from Kalman filter for data with both additive and innovation outliers

(a) Midpoint Count

0.2	2	**
0.3	9	*********
0.4	13	*************
0.5	8	********
0.6	16	****************
0.7	19	*******************
0.8	14	**************
0.9	12	************
1.0	4	****
1.1	2	**
1.2	1	*

(b) Midpoint Count

−0.8	1	*
−0.6	2	**
−0.4	9	*********
−0.2	14	**************
0.0	20	********************************
0.2	29	***
0.4	19	******************************
0.6	6	******

Figure 3.27 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

0.0	1	*
0.1	3	***
0.2	2	**
0.3	7	*******
0.4	7	*******
0.5	12	************
0.6	18	******************
0.7	17	*****************
0.8	14	**************
0.9	11	**********
1.0	5	*****
1.1	2	**
1.2	1	*

(b) Midpoint Count

−0.4	1	*
−0.3	1	*
−0.2	0	
−0.1	3	***
0.0	4	****
0.1	7	*******
0.2	6	******
0.3	10	**********
0.4	11	**********
0.5	16	**********
0.6	19	**********
0.7	9	*********
0.8	8	********
0.9	5	*****

Figure 3.28 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

1 observation below the first class

(a) Midpoint Count

–0.5	2	**
–0.4	1	*
–0.3	0	
–0.2	2	**
–0.1	3	***
0.0	4	****
0.1	11	**********
0.2	13	**********
0.3	20	**********
0.4	14	**********
0.5	10	**********
0.6	11	**********
0.7	4	****
0.8	2	**

(b) Midpoint Count

–0.6	1	*
–0.5	0	
–0.4	0	
–0.3	1	*
–0.2	4	****
–0.1	4	****
0.0	8	********
0.1	6	******
0.2	11	**********
0.3	12	**********
0.4	23	**********
0.5	12	**********
0.6	10	**********
0.7	2	**
0.8	3	***
0.9	1	*

Figure 3.29 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distribution for MLE of ln σ_ϵ when q = 1.5

Each * represents 2 observations

Midpoint	Count	
−14.0	1	*
−12.0	0	
−10.0	0	
−8.0	0	
−6.0	0	
−4.0	1	*
−2.0	0	
0.0	98	**********
2.0	1	*

Figure 3.30 MLE from the Kalman filter

Midpoint	Count	
−0.7	1	*
−0.6	0	
−0.5	0	
−0.4	3	***
−0.3	6	******
−0.2	9	*********
−0.1	12	**********
0.0	20	**********
0.1	36	**********
0.2	9	*********
0.3	4	****

Figure 3.31 MLE from the AGSF

Tracking of state (———) by AGSF (— —) and Kalman filter (– – –) with parameter estimation

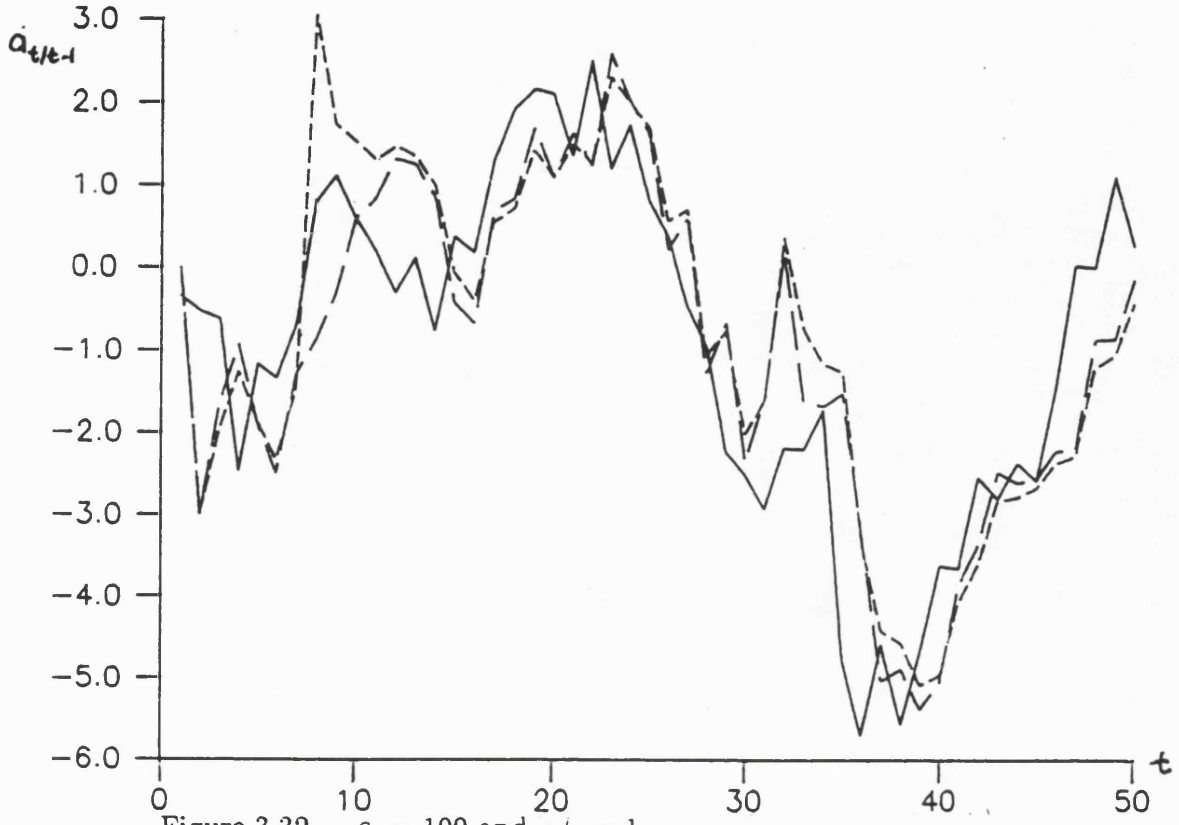


Figure 3.32 $g = 100$ and $b = 1$

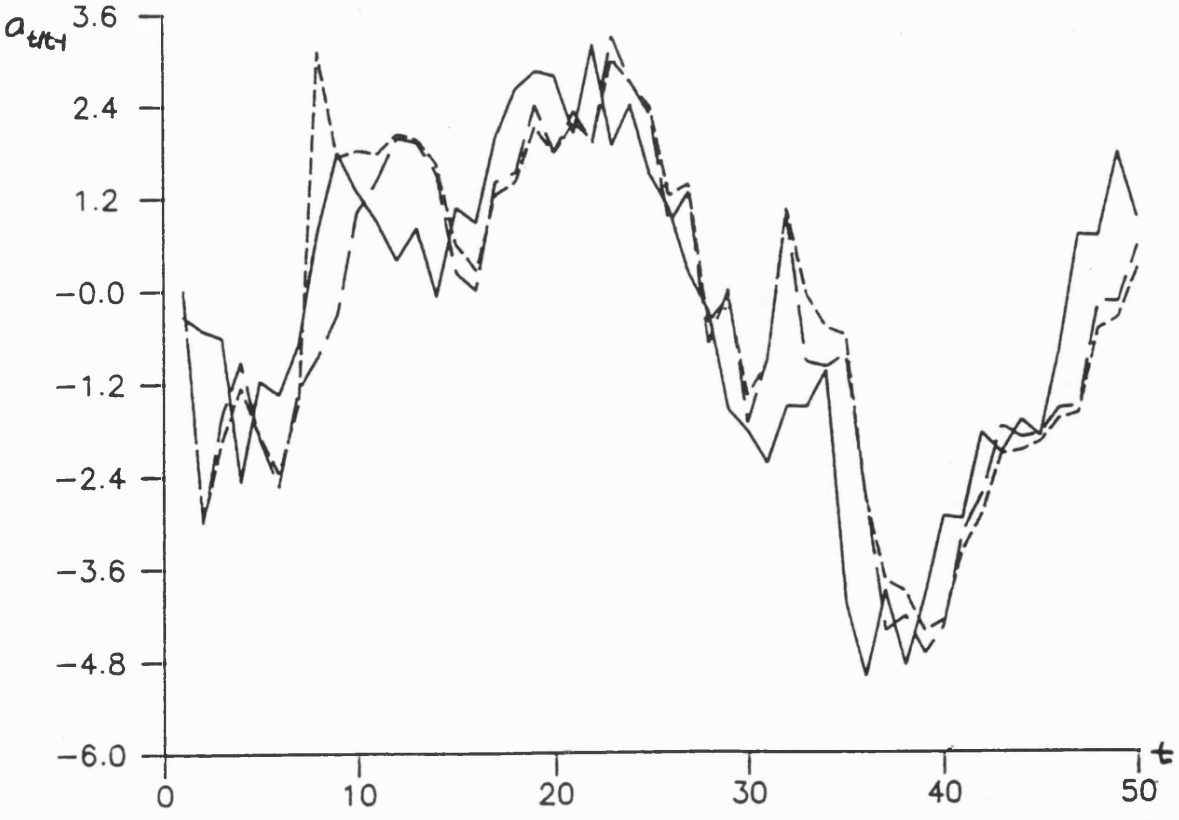


Figure 3.33 $g = 100$ and $b = 10$

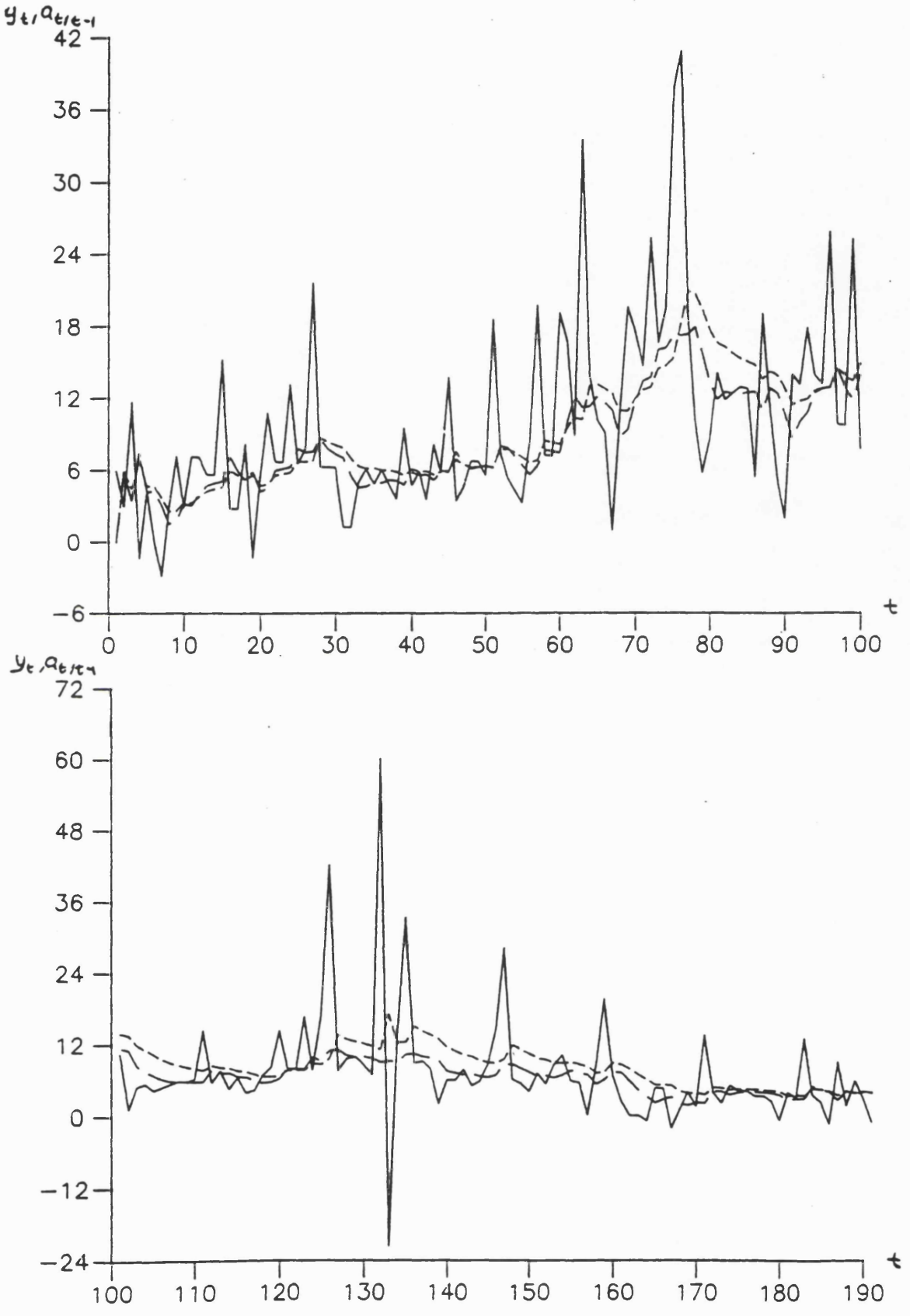


Figure 3.34 Plot of differenced UK price data (———) and one step ahead predictions from AGSF (— —) and the Kalman filter (– – –)

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln σ_η from Kalman filter for data with non–normal contamination

1 observation below the first class

(a) Midpoint Count

–0.3	1	*
–0.2	0	
–0.1	0	
0.0	3	***
0.1	2	**
0.2	13	**********
0.3	31	*******************************
0.4	26	**************************
0.5	13	**********
0.6	8	********
0.7	0	
0.8	1	*
0.9	1	*

(b) Midpoint Count

–0.6	2	**
–0.4	4	****
–0.2	23	**********************
0.0	34	**********************************
0.2	27	***************************
0.4	4	****
0.6	3	***
0.8	2	**
1.0	0	
1.2	0	
1.4	1	*

Figure 3.35 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

3 observations below the first class

(a) Midpoint Count

–0.2	1	*
–0.1	0	
0.0	5	*****
0.1	2	**
0.2	11	**********
0.3	29	*****************************
0.4	25	*************************
0.5	14	**************
0.6	8	*******
0.7	0	
0.8	1	*
0.9	1	*

1 observation above the last class

(b) Midpoint Count

–0.2	1	*
–0.1	3	***
0.0	3	***
0.1	9	*********
0.2	14	**************
0.3	13	*************
0.4	26	**************************
0.5	15	***************
0.6	7	*******
0.7	3	***
0.8	2	**
0.9	3	***

Figure 3.36 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

7 observations below the first class

(a) Midpoint Count

–0.6	1	*
–0.4	3	***
–0.2	14	**************
0.0	38	**************************************
0.2	26	**************************
0.4	7	*******
0.6	0	
0.8	2	**

6 observations above the last class

(b) Midpoint Count

–0.3	1	*
–0.2	4	****
–0.1	2	**
0.0	12	************
0.1	11	***********
0.2	19	*******************
0.3	23	***********************
0.4	15	***************
0.5	2	**
0.6	5	*****
0.7	1	*
0.8	0	
0.9	1	*
1.0	1	*

Figure 3.37 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) lnσ_η from AGSF for data with non–normal contamination

(a) Midpoint Count

–0.1	1	*
0.0	2	**
0.1	4	****
0.2	12	************
0.3	35	***********************************
0.4	32	********************************
0.5	12	************
0.6	2	**

(b) Midpoint Count

–1.4	2	**
–1.2	0	
–1.0	1	*
–0.8	0	
–0.6	5	*****
–0.4	7	*******
–0.2	28	****************************
0.0	33	*********************************
0.2	20	********************
0.4	4	****

Figure 3.38 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

–0.5	1	*
–0.4	0	
–0.3	0	
–0.2	0	
–0.1	3	***
0.0	1	*
0.1	3	***
0.2	16	****************
0.3	29	*****************************
0.4	27	***************************
0.5	14	**************
0.6	6	******

(b) Midpoint Count

−0.3	3	***
−0.2	0	
−0.1	5	*****
0.0	4	****
0.1	10	**********
0.2	17	*****************
0.3	17	*****************
0.4	24	************************
0.5	11	**********
0.6	5	*****
0.7	4	****

Figure 3.39 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

Each * represents two observations

(a) Midpoint Count

−1.6	1	*
−1.4	0	
−1.2	0	
−1.0	1	*
−0.8	0	
−0.6	3	**
−0.4	5	***
−0.2	14	*******
0.0	53	***************************
0.2	19	**********
0.4	2	*

(b) Midpoint Count

−1.4	1	*
−0.3	2	**
−0.2	4	****
−0.1	4	****
0.0	16	****************
0.1	16	****************
0.2	20	********************
0.3	18	******************
0.4	12	************
0.5	3	***
0.6	2	**

Figure 3.40 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Chapter Four

Robust Estimation of Scale

4.1 The Scale Estimation Problem

Consider the local level model ,

$$y_t = \alpha_t + \epsilon_t \tag{4.1.1a}$$

$$\alpha_t = \alpha_{t-1} + \eta_t \tag{4.1.1b}$$

where α_t is the state while ϵ_t and η_t are disturbance terms usually assumed to be normally distributed .

However , to model outliers in the system , we may allow the noise terms to take on mixture distributions . For example , if additive outliers are present , the density of ϵ_t can be chosen as

$$p(\epsilon_t) = (1-\gamma)N(0,\sigma_\epsilon^2) + \gamma\delta_\epsilon \tag{4.1.2}$$

where γ is the proportion of outliers and the contaminating component δ_ϵ has a distribution with longer tails than $N(0,\sigma_\epsilon^2)$.

In Chapters 2 and 3 , the main concern was robust sequential estimation of the state α_t . This corresponds to a location estimation problem in the i.i.d. case . The estimation of variances of the disturbance terms was carried out by the maximum likelihood estimation procedure . It was necessary to identify the level of contamination or the contaminating distribution by a grid search or by some information criterion such as the AIC .

However , in practice we may wish to treat the contaminating component as a nuisance parameter and not want to estimate it . Instead , the usual parameters of interest are the variances of the noise terms , σ_ϵ^2 and σ_η^2 , free from distortion by contamination . In this chapter , the aim is to produce estimates of these hyperparameters which are not sensitive to outliers .

4.2 M–estimates of Scale for i.i.d. case

First , let us consider a set of i.i.d. observations $x_1, x_2, \ldots, x_n$. Suppose each has density function $f_\theta(x_i)$ where θ is the scale parameter . Then , the scale model is given by

$$f_\theta(x_i) = \frac{1}{\theta} f(\frac{x_i}{\theta}) \tag{4.2.1}$$

Taking logarithms results in

$$\ln f_\theta(x_i) = -\ln \theta + \ln f(\frac{x_i}{\theta}) \tag{4.2.2}$$

and differentiation gives

$$\frac{\partial \ln f_\theta(x_i)}{\partial \theta} = -\frac{1}{\theta}\left[1 + \frac{f'(\frac{x_i}{\theta})}{f(\frac{x_i}{\theta})}\frac{x_i}{\theta}\right] \tag{4.2.3}$$

Scale invariance is achieved by multiplying the first derivative by the scale parameter , that is , the expression

$$\theta \frac{\partial \ln f_\theta(x_i)}{\partial \theta} = -\left[1 + \frac{f'(\frac{x_i}{\theta})}{f(\frac{x_i}{\theta})} \frac{x_i}{\theta}\right] \tag{4.2.4}$$

is independent of the unit of measurement . A maximum likelihood estimator (MLE) of the scale parameter is defined as the solution to

$$\sum_{i=1}^{n} \theta \frac{\partial \ln f_\theta(x_i)}{\partial \theta} = 0 \tag{4.2.5}$$

Hence , the influence function of a MLE takes the form

$$\Psi(y) = 1 + \frac{f'(y)}{f(y)} y \tag{4.2.6}$$

Such a influence function show that the MLE is not robust . Figure 4.1 gives a plot of the psi–function for normal observations . We note that its gross error sensitivity (GRE) is not bounded . It follows that an extreme observation could have unlimited influence on the estimator . To robustify the scale estimation procedure , we can use a generalised maximum likelihood estimator , alternatively known as M–estimator . We can modify the MLE by bounding its GRE . The psi–function in (4.2.6) is replaced by

$$\tilde{\Psi}(y) = \begin{cases} -b & \Psi^* < -b \\ \Psi(y) - c = \Psi^*(y) & \text{elsewhere} \\ b & \Psi^* > b \end{cases} \tag{4.2.7}$$

The new estimator of scale is obtained by equating $\sum_{i=1}^{n} \tilde{\Psi}(y_i)$ to zero . The modified score function is , in fact the original score function adjusted by c when the

magnitude of the derivative is smaller than a bound b . This is very like the MLE when the magnitude of all the derivatives is smaller than the bound b . But when this value exceeds the bound , the derivative is replaced by a constant , retaining its sign . In order to obtain Fisher consistency , the original psi–function $\Psi(y)$ has to be shifted by a constant c . Thus , the constant c is found by setting

$$E\left(\tilde{\Psi}(y)\right) = 0 \qquad (4.2.8)$$

where the expectation is taken over the distribution of y , usually normal . In this way , the new estimator is made comparable to the standard deviation at the normal distribution .

4.3 Scale Model and Sensitivity Curves for Time Dependent Observations

Consider a time series $y_1, y_2, \ldots, y_T$. Although the observations are now time dependent , their joint likelihood can be broken up into conditional densities $p(y_t/Y_{t-1})$. Hence , the probability density function in the i.i.d case $f_\theta(x_i)$ is now replaced by the conditional likelihood $p(y_t/Y_{t-1})$. The latter is a function of the one–step ahead state prediction error v_t and its scale f_t . Unlike the previous section , both the location $a_{t/t-1}$ and the scale f_t are now dependent on the parameters of interest . To make the situation more comparable to the i.i.d. case , reparameterisation is appropriate . We would like to estimate the standard deviation of the measurement noise term σ_ϵ and the ratio of the standard deviation of the two noise terms $\sqrt{q} = \frac{\sigma_\eta}{\sigma_\epsilon}$.

4.3.1 Sensitivity curve for time dependent observations

To gain a better understanding of the problem , we shall first see how sensitive the MLE of $\sqrt{q}$, which is denoted by $\hat{\sqrt{q}}$, is towards outliers . In other words , we want to investigate the sensitivity curve of $\hat{\sqrt{q}}$, assuming σ_ϵ given . According to the definition of a sensitivity curve , the change in the ML estimate of $\sqrt{q}$ is measured when an additional observation (x_n) is included in the sample of size n–1 . Unlike the i.i.d. situation where the position of this new observation in the sample is not important , it makes a difference as to where we place the new observation in the series . For simplicity , let us consider the new observation being added to the end of the series . This corresponds to the notion of conditional influence function discussed in Kunsch (1984) and is appropriate in the prediction error decomposition approach . The sensitivity curve denoted by SC is the difference in the estimates taken as a proportion of change in the sample size and is a function of the additional observation. Thus , we have

$$SC(x_n) = \frac{\hat{\sqrt{q}}(x_1,x_2,\ldots,x_{n-1}) - \hat{\sqrt{q}}(x_1,x_2,\ldots,x_n)}{\frac{1}{n-1}} \qquad (4.3.1)$$

Two series were generated from the local level model given in (4.1.1) with normally distributed disturbances defined as $\epsilon_t \sim N(0,2)$ and $\eta_t \sim N(0,1)$, i.e. q = 0.5 . The first is of length 300 and the second has length 400 . Plots of the sensitivity curves for these two series are found in figures 4.2a and 4.2b . The functions turn out to be similar to that of a scale variable in the i.i.d. case . They are both quadratic and symmetric about zero which means that deviation of the new observation from the actual observation in either direction will influence the MLE of

$\sqrt{q}$ in a similar way . This should be the case since $\sqrt{q}$ is a symmetric parameter . When the additional observation assumes an extreme value of either sign , it becomes an outlier , thereby increasing the estimate for system noise standard deviation σ_η since the measurement noise standard deviation σ_ϵ is fixed and known . A resulting increase in the new MLE of $\sqrt{q}$ leads to a negative value for the sensitivity curve . Hence , the functions are negative and downward sloping . We note that there is stability in the shape of the curves for the two series of different lengths . We next generate another series of length three hundred from the same model but with different hyperparameter values , namely $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 5$ i.e. $q = 5$. The sensitivity curve for this series with bigger $\sqrt{q}$ value , see figure 4.2c , has a shape similar to the above functions . However , we see that the estimator is more sensitive to outliers in this case .

4.3.2 The Kalman Filter and the Likelihood Function

With reparameterisation , the local level model can be set up as

$$y = \alpha_t + \epsilon_t \qquad \mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2 \tag{4.3.2a}$$

$$\alpha_t = \alpha_{t-1} + \eta_t \qquad \mathrm{Var}(\eta_t) = q\sigma_\epsilon^2 \tag{4.3.2b}$$

Let $\underset{\sim}{\theta}$ be the vector of parameters , i.e. , $\underset{\sim}{\theta} = (\sigma_\epsilon, \sqrt{q})'$. It is not necessary to involve the hyperparameter σ_ϵ^2 in the recursive equations of the Kalman filter . Thus , we have for

(i) Prediction

$$a_{t/t-1} = a_{t-1} \tag{4.3.3a}$$

$$P_{t/t-1} = P_{t-1} + q \tag{4.3.3b}$$

(ii) Updating

$$a_t = a_{t/t-1} + P_{t/t-1} v_t / f_t \tag{4.3.3c}$$

$$P_t = P_{t/t-1} - P^2_{t/t-1} / f_t \tag{4.3.3d}$$

where

$$v_t = y_t - a_{t/t-1} \tag{4.3.3e}$$

$$f_t = P_{t/t-1} + 1 \tag{4.3.3f}$$

Under the normality assumption of the disturbance terms , the conditional density of each observation $p(y_t/Y_{t-1})$ is normal with mean $a_{t/t-1}$ and variance $\sigma^2_\epsilon f_t$. Hence, the log likelihood function is given by

$$\sum_{t=1}^{T} \ln p(y_t/Y_{t-1}) = \sum_{t=1}^{T} \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln f_t \sigma^2_\epsilon - \frac{1}{2} \frac{(y_t - a_{t/t-1})^2}{f_t \sigma^2_\epsilon} \right] \tag{4.3.4}$$

This is is often replaced by the following objective function

$$S(\underset{\sim}{\theta}) = \sum_{t=1}^{T} \ln f_t \sigma^2_\epsilon + \sum_{t=1}^{T} \frac{(y_t - a_{t/t-1})^2}{f_t \sigma^2_\epsilon} \tag{4.3.5}$$

In order to perform optimisation over one instead of two parameters , σ^2_ϵ is usually concentrated out of the likelihood function . This is carried out by replacing σ^2_ϵ by its MLE which is

$$\hat{\sigma}_\epsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{v_t^2}{f_t} \tag{4.3.6}$$

However , after substituting (4.3.6) into (4.3.5) , the resulting form of the objective function will have a derivative which is quite complicated so that modifications done in the i.i.d. case cannot be carried over easily . Therefore , to simplify the expression for the first derivative and to make it more comparable with the i.i.d. case , both parameters σ_ϵ and $\sqrt{q}$ will be estimated simultaneously .

4.3.3 Derivatives of the Likelihood function

It is clear from (4.3.5) that the first derivative of the objective function will involve the derivative of f_t and $a_{t/t-1}$ and the Kalman filter equations in (4.3.3) show that these variables are dependent on the parameter $\sqrt{q}$. The first derivative of f_t and $a_{t/t-1}$ with respect to (w.r.t.) $\sqrt{q}$ can be evaluated by running another set of recursions in parallel with the Kalman filter . Corresponding to the set of equations in (4.3.3) , they are

$$\frac{\partial\, a_{t/t-1}}{\partial\sqrt{q}} = \frac{\partial\, a_{t-1}}{\partial\sqrt{q}} \tag{4.3.7a}$$

$$\frac{\partial\, P_{t/t-1}}{\partial\sqrt{q}} = \frac{\partial\, P_{t-1}}{\partial\sqrt{q}} + 2\sqrt{q} \tag{4.3.7b}$$

$$\frac{\partial\, a_t}{\partial\sqrt{q}} = \frac{\partial\, a_{t/t-1}}{\partial\sqrt{q}}\left[1 - \frac{P_{t/t-1}}{f_t}\right] + \frac{v_t}{f_t}\left[\frac{\partial\, P_{t/t-1}}{\partial\sqrt{q}} - \frac{P_{t/t-1}}{f_t}\frac{\partial\, f_t}{\partial\sqrt{q}}\right] \tag{4.3.7c}$$

$$\frac{\partial P_t}{\partial \sqrt{q}} = \frac{\partial P_{t/t-1}}{\partial \sqrt{q}} \left[1 - 2\frac{P_{t/t-1}}{f_t}\right] + \left[\frac{P_{t/t-1}}{f_t}\right]^2 \frac{\partial f_t}{\partial \sqrt{q}} \tag{4.3.7d}$$

where

$$\frac{\partial v_t}{\partial \sqrt{q}} = -\frac{\partial a_{t/t-1}}{\partial \sqrt{q}} \tag{4.3.7e}$$

$$\frac{\partial f_t}{\partial \sqrt{q}} = \frac{\partial P_{t/t-1}}{\partial \sqrt{q}} \tag{4.3.7f}$$

On the other hand , the recursive relations in the Kalman filter (4.3.3) do not involve the parameter σ_ϵ . This implies that the derivative of f_t and $a_{t/t-1}$ w.r.t. σ_ϵ are zero. The actual expressions for the derivative of the objective function will be given in the next section 4.4 .

4.3.4 Steady State

Another representation of the Kalman filter can be achieved by combining the prediction equations with the updating equations , namely (4.3.3a) with (4.3.3c) and (4.3.3b) with (4.3.3d) . Hence , the recursive equations can be written as

$$a_t = a_{t-1} + P_{t/t-1} v_t / f_t \tag{4.3.8a}$$

$$\Rightarrow \quad a_t = \frac{1}{f_t} a_{t-1} + \frac{f_t^{-1}}{f_t} y_t \tag{4.3.8b}$$

and

$$P_{t+1/t} = P_{t/t-1} - P^2_{t/t-1}/f_t + q \tag{4.3.9}$$

where v_t and f_t are given in (4.3.3e) and (4.3.3f) respectively .

In order to obtain an explicit expression for the psi–functions like (4.2.7) in the i.i.d. case , we shall consider observations after the steady state is reached . When the Kalman filter reaches steady state , the variance of the state becomes time invariant . Thus , (4.3.9) becomes

$$\bar{P} = \bar{P} - \bar{P}^2/f + q \qquad (4.3.10a)$$

where $f = \bar{P} + 1$ (4.3.10b)

$$\Rightarrow \quad q = \bar{P}^2/(\bar{P} + 1) \qquad (4.3.10c)$$

$\bar{P}$ denotes the error covariance matrix after steady state is reached ,
This is the one–one functional relation between the parameter of interest q and $\bar{P}$.

The steady state recursions for the state is

$$a_t = \frac{1}{f} a_{t-1} + \frac{f-1}{f} y_t \qquad (4.3.11)$$

Repeated substitution gives

$$a_t = \sum_{i=1}^{t-m} \frac{f-1}{f^i} y_{t-i+1} + \frac{1}{f^{t-m}} a_m \qquad (4.3.12)$$

where steady state starts from the (m+1)th observation . Suppose only steady state observations are used , then , the last term in the above expression is dropped . This gives us

$$a_t \approx \sum_{i=1}^{t-m} \frac{f-1}{f^i} y_{t-i+1} \qquad (4.3.13)$$

Since $f \geq 1$, this is not a bad approximation especially when t is considerably greater than m . Differentiating a_t w.r.t. f results in

$$\frac{\partial a_t}{\partial f} \approx \sum_{i=1}^{t-m} \frac{(1-i)f + i}{f^{i+1}} y_{t-i+1} \tag{4.3.14}$$

Excluding all observations before steady state is reached , the objective function in (4.3.5) becomes

$$S(\underset{\sim}{\theta}) \approx \sum_{t=m+1}^{T} \ln f\sigma_\epsilon^2 + \sum_{t=m+1}^{T} \frac{(y_t - a_{t-1})^2}{f\sigma_\epsilon^2} \tag{4.3.15}$$

where f and a_{t-1} are defined in (4.3.10b) and (4.3.13) respectively .

4.4 M–estimators for Time Dependent Observations

4.4.1 Psi–functions for Parameters of Interest

Suppose the estimation of the two parameters , σ_ϵ and $\sqrt{q}$, is carried out separately . Let us first consider the estimation of σ_ϵ when $\sqrt{q}$ is given . The differentiation of the tth component of the objective function (4.3.5) w.r.t. σ_ϵ is

$$\frac{\partial S(\sigma_\epsilon)}{\partial \sigma_\epsilon} = \frac{2}{\sigma_\epsilon}\left[1 - \frac{v_t^2}{f_t \sigma_\epsilon^2}\right] \tag{4.4.1}$$

To ensure that the estimator for σ_ϵ does not depend on the unit of measurement , the psi–function is given by

$$\Psi(\sigma_\epsilon,t) = \sigma_\epsilon \frac{\partial\ S(\sigma_\epsilon)}{\partial\ \sigma_\epsilon} \tag{4.4.2}$$

Considering only steady state observations and substituting (4.4.1) into (4.4.2) , we have

$$\Psi(\sigma_\epsilon,t) \approx 2\left[1 - \frac{v_t^2}{f\sigma_\epsilon^2}\right] \tag{4.4.3}$$

It follows that a M–estimator of σ_ϵ with bounded gross error sensitivity is the root of the following function

$$\sum_{t=1}^{T} \tilde{\Psi}(\sigma_\epsilon,t) \tag{4.4.4a}$$

where

$$\tilde{\Psi}(\sigma_\epsilon,t) = \begin{cases} -b_1 & \Psi^* < -b_1 \\ \Psi(\sigma_\epsilon,t) - c_1 = \Psi^*(\sigma_\epsilon,t) & \text{elsewhere} \\ b_1 & \Psi^* > b_1 \end{cases} \tag{4.4.4b}$$

This estimator has a GRE which is related to b_1 and c_1 is the shift in the psi–function needed to obtain Fisher consistency .

Next , we consider the estimation of $\sqrt{q}$ assuming σ_ϵ to be known . The derivative of the tth term of the objective function in (4.3.5) w.r.t. $\sqrt{q}$ is

$$\frac{\partial\, S(\sqrt{q})}{\partial\, \sqrt{q}} = \frac{1}{f_t}\frac{\partial f_t}{\partial\sqrt{q}} + \frac{1}{\sigma_\epsilon^2}\left[\frac{2v_t}{f_t}\frac{\partial\, v_t}{\partial\sqrt{q}} - \left[\frac{v_t}{f_t}\right]^2 \frac{\partial f_t}{\partial\sqrt{q}}\right] \tag{4.4.5}$$

Since $v_t = y_t - a_{t-1}$, it follows that

$$\frac{\partial\, v_t}{\partial\sqrt{q}} = -\frac{\partial a_{t-1}}{\partial\sqrt{q}}$$

$$= -\frac{\partial a_{t-1}}{\partial\, f_t}\frac{\partial f_t}{\partial\sqrt{q}} \tag{4.4.6}$$

Substituting (4.4.6) into (4.4.5) gives

$$\frac{\partial\, S(\sqrt{q})}{\partial\sqrt{q}} = \frac{1}{f_t}\frac{\partial f_t}{\partial\sqrt{q}} + \frac{1}{\sigma_\epsilon^2}\frac{1}{f_t}\frac{\partial f_t}{\partial\sqrt{q}}\left[-2v_t\frac{\partial a_{t-1}}{\partial\, f_t} - \frac{v_t^2}{f_t}\right] \tag{4.4.7}$$

Consider a psi–function which is this differential scaled by the derivative of the ln f , i.e. ,

$$\Psi(\sqrt{q},t) = \frac{\partial\, S(\sqrt{q})}{\partial\sqrt{q}} \Big/ \frac{1}{f_t}\frac{\partial f_t}{\partial\sqrt{q}} \tag{4.4.8}$$

Since the steady state is reached quickly , the scaling term is constant at $\frac{1}{f}\frac{\partial f}{\partial\sqrt{q}}$ most of the time . Substituting (4.4.7) into (4.4.8) results in

$$\Psi(\sqrt{q},t) = 1 + \frac{1}{\sigma_\epsilon^2}\left[-2v_t\frac{\partial a_{t-1}}{\partial\, f_t} - \frac{v_t^2}{f_t}\right] \tag{4.4.9}$$

Considering steady state observations only and substituting (4.3.14) into (4.4.9) gives

$$\Psi(\sqrt{q},t) \approx 1 + \frac{1}{\sigma_\epsilon^2}\left[-2v_t \sum_{i=1}^{t-m} \frac{(1-i)f + i}{f^{i+1}} y_{t-i+1} - \frac{v_t^2}{f}\right] \tag{4.4.10}$$

We note that this psi–function is independent of the unit of measurement . Hence , a robust estimate of $\sqrt{q}$ is the root of the following function

$$\sum_{t=1}^{T} \tilde{\Psi}(\sqrt{q},t) \tag{4.4.11a}$$

where

$$\tilde{\Psi}(\sqrt{q},t) = \begin{cases} -b_2 & \Psi^* < -b_2 \\ \Psi(\sqrt{q},t) - c_2 = \Psi^*(\sqrt{q},t) & \text{elsewhere} \\ b_2 & \Psi^* > b_2 \end{cases} \tag{4.4.11b}$$

This M–estimate of $\sqrt{q}$ has GRE bounded at b_2 and c_2 is the shift in the psi–function needed to attain Fisher consistency .

All the observations , not only the steady state ones are used in the computation of the M–estimates . We run the two sets of recursions , namely the Kalman filter in section 4.3.2 and its derivatives in section 4.3.3 , on the series in question . Instead of evaluating the expressions (4.4.3) and (4.4.10) directly , we can use (4.4.2) and (4.4.8) to evaluate the psi–functions at each stage . Since the score functions are truncated , the estimates are no longer consistent . Some adjustments by the terms c_1 and c_2 in (4.4.4b) and (4.4.11b) is necessary to achieve Fisher consistency . Having fixed the bounds b_1 and b_2 as constants , the values of c_1 and c_2 are determined by letting the mean of the psi–functions $\tilde{\Psi}(\sigma_\epsilon,t)$ and $\tilde{\Psi}(\sqrt{q},t)$ be

zero , i.e. ,

$$E(\tilde{\Psi}(\sigma_\epsilon,t)) = 0 \qquad (4.4.12a)$$

and $$E(\tilde{\Psi}(\sqrt{q},t)) = 0 \qquad (4.4.12b)$$

where the expectation is taken over the distribution of y .

The terms c_1 and c_2 can be found by Monte Carlo methods described later in section 4.5.1 .It turns out that both c_1 and c_2 are functions of the parameter $\sqrt{q}$.

4.4.2 Two–dimensional Huber Function

The purpose of this section is to obtain M–estimates of the parameters σ_ϵ and $\sqrt{q}$ simultaneously . That is , we are looking at a generalisation of the robust estimation procedure for a single parameter to a multi–dimensional context . A discussion of the generalised Huber function can be found in Hampel *et al* (1986). Let the two psi–functions $\Psi^*(\sigma_\epsilon,t)$ and $\Psi^*(\sqrt{q},t)$ from (4.4.4b) and (4.4.11b) respectively form a vector $\underset{\sim}{z}$. The magnitude of this vector is measured by its norm which is

$$\|\underset{\sim}{z}\| = \sqrt{\Psi^*(\sigma_\epsilon,t)^2 + \Psi^*(\sqrt{q},t)^2} \qquad (4.4.13)$$

The bound of this two–dimensional psi–function , $\underset{\sim}{z}$, is now the surface of a sphere with radius b . No modification of the estimates are needed if the norm of $\underset{\sim}{z}$ is smaller than the bound b . But when the length of $\underset{\sim}{z}$ exceeds this bound , $\underset{\sim}{z}$ is replaced by its nearest point on the sphere . This two–dimensional Huber function can be written as

$$h(\underset{\sim}{z},t) = \underset{\sim}{z} \min\left[1 , \frac{b}{\|\underset{\sim}{z}\|}\right] \qquad (4.4.14)$$

and a plot of this function is shown in figure 4.3

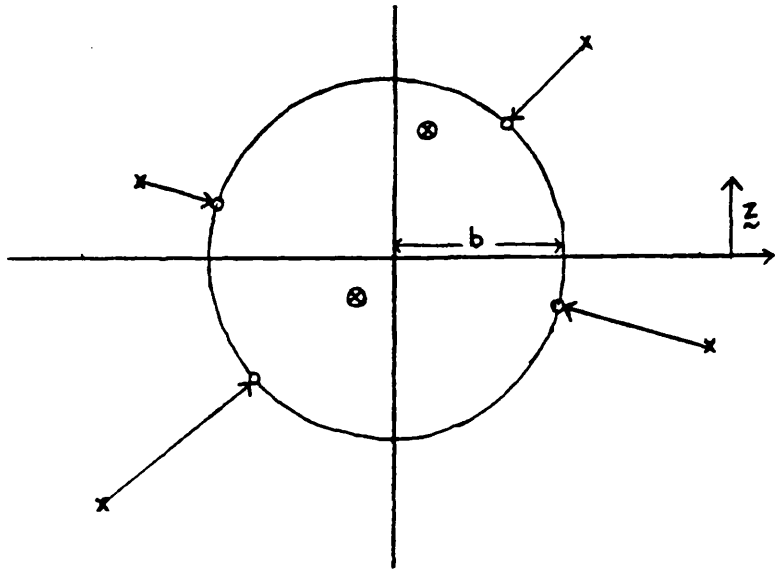


Figure 4.3 Two dimensional Huber function $\underset{\sim}{z}$ –> h($\underset{\sim}{z}$,t)

Suppose the parameters σ_ϵ and $\sqrt{q}$ have quite different order of magnitude . In the above estimation procedure , when either one of the psi–functions is zero , the other estimator will have the same cut–off , b . Such symmetric treatment of the the two parameters will then be inappropriate . Hence , weights should be attached to the psi–functions to increase efficiency in estimation . Leaving the coefficient of the psi–function $\Psi^*(\sqrt{q},t)$ as one , a coefficient d is attached to the psi–function $\Psi^*(\sigma_\epsilon,t)$. Thus , the vector $\underset{\sim}{z}$ in the two–dimensional Huber function in (4.4.14) is now defined as

$$\underset{\sim}{z} = \begin{bmatrix} d\Psi^*(\sigma_\epsilon,t) \\ \Psi^*(\sqrt{q},t) \end{bmatrix} \qquad (4.4.15)$$

The value of d is chosen to make the variance of the psi–functions equal , i.e. ,

$$E(d\Psi^*(\sigma_\epsilon,t)^2) = E(\Psi^*(\sqrt{q},t)^2) \qquad (4.4.16)$$

From the initial estimates of scale , described in Chapter 3 , we can roughly gauge their order of magnitude . If these turn out to be very different , we can try to set a value for d so that more weight is given to the smaller parameter . Recalibration to obtain the corresponding values for c_1 and c_2 is necessary . For our investigations below , d is kept at the value of one .

4.4.3 Starting Values

Once we have found the functional relation of c_1 and c_2 in terms of the parameters σ_ϵ and $\sqrt{q}$, M–estimates of these parameters can be obtained by solving for the roots of the Huber function in (4.4.14) . Since this is a two–dimensional search , it is important to have good starting values . These are usually resistant estimates derived from the sample variance and covariance of a differenced series , see Chapter 3 . Alternatively , we can use estimates of hyperparameters obtained by running another filter such as the AGSF on the series as starting values .

4.5 Simulation results

4.5.1 Calibration

With reference to equations (4.4.4b) and (4.4.11b) , we see that the

M–estimators need adjustment in order to achieve Fisher consistency . To perform the calibration of these estimators , we express the two–dimensional Huber function in (4.4.14) as

$$h(\underset{\sim}{z},t) = \begin{bmatrix} h_1(\sigma_\epsilon,t) \\ h_2(\sqrt{q},t) \end{bmatrix} \tag{4.5.1}$$

Letting the GRE be 0.5 , the values of c_1 and c_2 are found by solving the following respective equations

$$\sum_{t=1}^{T} h_1(\sigma_\epsilon,t) = 0 \tag{4.5.2a}$$

$$\sum_{t=1}^{T} h_2(\sqrt{q},t) = 0 \tag{4.5.2b}$$

A root–finding routine from the NAG library C05NBF is used to compute c_1 and c_2 for each generated series .

Unlike the i.i.d.situation , the terms c_1 and c_2 are no longer constants but are likely to be functions of one or both parameters of interest . Hence , the values of c_1 and c_2 are obtained over a grid of values for σ_ϵ and $\sqrt{q}$. The values used for the scale of the observation noise are 1.00 , 1.75 and 2.25 while the parameter $\sqrt{q}$ ranges from 0.1 to 10. For each pair of values of the parameters , one hundred series each of length five hundred are generated from the local level model . A long series ensures that its hyperparameter values are actually the same as those used to generate it . One hundred replications are used for each case to allow for sampling variability in the computed values of c_1 and c_2 . This means that for each set of parameter values ,

we have the sampling distributions of c_1 and c_2 . Only the means of the distributions are used in the calibration and they are tabulated below .

TABLE 4.1 CALIBRATION RESULTS FOR M–ESTIMATORS

$\sqrt{q}$	σ_ϵ	c_1	c_2
0.1	1.00	1.2893	0.7124
	1.75	1.2907	0.6822
	2.25	1.2980	0.7081
0.3162	1.00	1.2255	0.6426
	1.75	1.2271	0.6729
	2.25	1.2378	0.6569
0.7071	1.00	1.1767	0.6126
	1.75	1.1666	0.6082
	2.25	1.1579	0.6034
1.0	1.00	1.1039	0.5658
	1.75	1.1260	0.5731
	2.25	1.0945	0.5612
2.0	1.00	1.0678	0.5390
	1.75	1.0606	0.5343
	2.25	1.0662	0.5369
3.1623	1.00	1.0608	0.5324
	1.75	1.0628	0.5320
	2.25	1.0664	0.5355
10.0	1.00	1.0568	0.5284
	1.75	1.0590	0.5296
	2.25	1.0543	0.5273

A statistical package Minitab is used to analyse the results in Table 4.1 in order to find expressions for c_1 and c_2 in terms of the parameter σ_ϵ and $\sqrt{q}$. It turns out that neither c_1 nor c_2 depends on the parameters σ_ϵ , see figures 4.4a and 4.5a . On the other hand , figures 4.4b and 4.5b reveal that both c_1 and c_2 have a well–defined relation with the parameter $\sqrt{q}$. Using power transformations , we can write the functions as follows .

$$c_1 = 0.994 + 0.111/(\sqrt{q})^{.3} \qquad R^2 = 97.0\% \tag{4.5.3a}$$

$$c_2 = 0.483 + 0.081/(\sqrt{q})^{.3} \qquad R^2 = 96.3\% \tag{4.5.3b}$$

Note that the high R^2 values indicate a good fit for each case . A comparison of the fitted and actual values of c_1 and c_2 , see figures 4.4c and 4.5c , confirms the above relations .

In order to see whether the length of the series has any effect on the calibration terms , we repeated the calibration procedure for shorter series of length two hundred. The hyperparameter values of the series are not necessarily the same as those values used to generate them because they are not long enough . Hence , we expect more variability in the calibration terms and the number of replications used for each set of hyperparameter values is increased to two hundred to cope with this problem . Otherwise , figures 4.6a and 4.6b reveal that the functions c_1 and c_2 in terms of $\sqrt{q}$ are very similar to those given in (4.5.3) . We , therefore , conclude that the calibration terms are invariant to the length of the series and that the relations in (4.5.3) can be applied to series of different lengths .

4.5.2 Performance of M–estimators

We can investigate the performance of the M–estimators of scale for different hyperparameter values by simulation . The three cases considered are

(a) $\sigma_\epsilon^2 = 2$, $\sigma_\eta^2 = 1$, i.e. , q = 0.5
(b) $\sigma_\epsilon^2 = 2$, $\sigma_\eta^2 = 2$, i.e. , q = 1.0 and
(c) $\sigma_\epsilon^2 = 1$, $\sigma_\eta^2 = 1.5$, i.e. , q = 1.5

For each case , one hundred random series each of length a hundred are generated from the local level model . Later in this section , we see that one hundred replications used in each simulation is enough to reveal differences amongst the various estimators . As discussed in Chapter 3 , the parameter σ_ϵ^2 is sometimes estimated as zero when the value of q is big . When the ML estimation procedure is carried out with clean data , this happened at replication 57 . Hence , to remove this effect , replication 57 is discarded whenever we investigate case (c) . We start off with clean data where both disturbances are generated from the normal distributions $N(0,\sigma_\epsilon^2)$ and $N(0,\sigma_\eta^2)$. A root–finding routine from the NAG library C05NBF is used to estimate the parameters σ_ϵ and $\sqrt{q}$ for each series . The estimates are values which equate the two–dimensional Huber function in (4.4.14) to zero . Note that c_1 and c_2 are replaced by the functions in (4.5.3) found during the calibration process . The root–finding routine sometimes does not converge to the global minimum of the negative log–likelihood function . To understand this phenomenon , we look at the contour map for the objective function of one such case , see figure 4.7 . Although the objective function has a well–defined minimum , see figure 4.7a , it also has a long plateau to the right , see figure 4.7b . When the routine is searching over this region, it becomes difficult to distinguish between big or small values of ln $\sqrt{q}$. To

overcome this problem , different starting values especially those which avoid the flat region are used when non–convergence is indicated by the root–finding routine . We define breakdown cases to be those which produce estimates which are very far from the minimum .

We judge the performance of the M–estimators by looking at the summary statistics of the estimates taken over the one hundred replications . This number of replications is enough to reveal differences between the different types of estimators . We perform a logarithmic transformation on the hyperparameters in order to facilitate the comparison amongst the three cases . Hence , we estimate the logarithms of the scales , i.e. ln σ_ϵ and ln $\sqrt{q}$, in the three sets of simulations . The sampling distributions of the two estimators for the three cases are found in figures 4.8 to 4.10 . Table 4.2 below shows the bias and mean square error (MSE) of the M–estimators for the different cases . As in Chapter 3 , if there is at least one breakdown in any set of simulations , another set of summary statistics which are more resistant is computed . These are the bias , bias* , evaluated using the median and the mean square error , MSE* , computed using the pseudovariance . The number of cases which are excluded in the computation of these statistics , M , are also indicated when presenting this alternative set of summary statistics .

TABLE 4.2 SUMMARY STATISTICS FOR M–ESTIMATORS (clean data)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	–0.0511	0.0489
(b)	0.3466	–0.0858	0.0848
(c)	0.0	–0.1110	0.1305
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.0487	0.2375
(b)	0.0	0.0087	0.2616
(c)	0.2027	0.0381	0.2947

To better judge the performance of the M–estimators , we shall compare them with the MLEs produced from the Kalman filter . The MLE of the hyperparameters can be found by equating the psi–functions in (4.4.2) and (4.4.8) directly to zero . No calibration terms are needed here since the psi–functions are not truncated . We repeat the above simulations , finding ML estimates for the same series generated before . The sampling distributions of the two MLEs for the three cases are given in figures 4.11 to 4.13 . Their summary statistics are given in Table 4.3 below .

TABLE 4.3 SUMMARY STATISTICS FOR MLE (clean data)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	–0.016	0.0128
(b)	0.3466	–0.0264	0.0221
(c)	0.0	–0.0391	0.0361
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.0017	0.0917
(b)	0.0	0.0126	0.0858
(c)	0.2027	0.0269	0.1026

From the above tables , we observe that for the parameter $\ln \sigma_\epsilon$, both the bias and mean square error of the sampling distributions for the MLEs are significantly smaller than those of the M–estimators . On the other hand , there is no significant difference in the bias for the two types of estimators of $\ln \sqrt{q}$, although there is a larger difference in their mean square errors . We also see that the sampling distributions of the MLEs are much tighter than those of the M–estimators. Hence , the MLEs perform better than the M–estimators which is expected in a situation where there are no outliers since the MLEs are best estimators under normality .

We next compare the performance of the M–estimators with the MLEs in the case where there is contamination in the series . To allow for additive outliers only in the series , the observation noise term is generated from a contaminated normal distribution , i.e. ,

$$p(\epsilon_t) = (1-\gamma)N(0,\sigma_\epsilon^2) + \gamma N(0,k\sigma_\epsilon^2) \tag{4.5.4}$$

where the contamination factor k corresponds to g in chapter three.
while the disturbance term in the transition equation is generated from a normal distribution $N(0,\sigma_\eta^2)$. The percentage of outliers is fixed at 5% in the simulations for convenience . In order to maintain roughly the same the level of contamination in all three cases , we consider the reduced form of the local level model . The differenced series and its variance are given by

$$\Delta y_t = \eta_t + \epsilon_t - \epsilon_{t-1} \tag{4.5.5}$$

$$\text{Var}(\Delta y_t) = q\sigma_\epsilon^2 + 2\,\text{Var}(\epsilon_t) \tag{4.5.6}$$

Let us denote the variance of the differenced series with and without contamination as $\text{Var}_c(\Delta y_t)$ and $\text{Var}_n(\Delta y_t)$ respectively . Hence , we have

$$\text{Var}_n(\Delta y_t) = (q+2)\sigma_\epsilon^2 \tag{4.5.7}$$

$$\begin{aligned}\text{Var}_c(\Delta y_t) &= q\sigma_\epsilon^2 + 2\,(0.95 + .05k)\sigma_\epsilon^2 \\ &= (1.9 + 0.1k + q)\sigma_\epsilon^2\end{aligned} \tag{4.5.8}$$

The ratio of these two variances , denoted by r , is

$$r = \frac{\text{Var}_c(\Delta y_t)}{\text{Var}_n(\Delta y_t)}$$

$$= \frac{1.9 + 0.1k + q}{2 + q} \tag{4.5.9}$$

To keep a comparable level of contamination in all three cases of different hyperparameter values , we just have to fix the ratio r at the same constant in each case . Let the factor k be $\frac{(20)}{0.5}$ in case (a) . Then ,

$$r = \frac{1.9 + (0.1)(20) + 0.5}{2 + 0.5}$$
$$= 1.76 \qquad (4.5.10)$$

Keeping the ratio fixed at 1.76 , the following relation is derived from (4.5.9)

$$k = 16.2 + 7.6q \qquad (4.5.11)$$

Hence , the corresponding factors of contamination in cases(b) and (c) work out to be 23.8 and 27.6 respectively .

The above simulations to evaluate the M–estimates and ML estimates for series with different hyperparameter values are repeated with contaminated data . The levels of contamination in the three cases are (a) $\gamma = 5\%$, $k = 20$; (b) $\gamma = 5\%$, $k = 25$ and (c) $\gamma = 5\%$, $k = 30$. Sampling distributions of M–estimators are found in figures 4.14 to 4.16 while those of the MLEs are given in figures 4.17 to 4.19. Tables 4.4 and 4.5 contain the summary statistics derived from the simulation results.

TABLE 4.4 SUMMARY STATISTICS FOR M–ESTIMATORS (γ=5%)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	0.0733	0.0481
(b)	0.3466	0.0692	0.0836
(c)	0.0	0.0864	0.0990
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.1818	0.2779
(b)	0.0	–0.1322	0.2783
(c)	0.2027	–0.1417	0.2585

TABLE 4.5 SUMMARY STATISTICS FOR MLE (γ=5%)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	0.2999	0.1357
(b)	0.3466	0.3457	0.1808
(c)	0.0	0.3905	0.2273
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.4057	0.2936
(b)	0.0	–0.4233	0.3142
(c)	0.2027	–0.4605	0.3575

The simulation results show that the MLEs are more affected by the outliers

than the M–estimators are . There is a considerable shift in the sampling distributions of the MLEs resulting in significantly bigger biases . The mean square errors for the MLEs are also smaller than those for the M–estimators . This difference is more significant for the parameter ln $\sqrt{q}$. Such poor performance of the MLEs is because we are performing maximum likelihood estimation when there is a misspecification in the model , see White (1982) . Such estimators derived from optimising an incorrectly assumed Gaussian likelihood are known as quasi–MLEs . Suppose that the fourth moment of the disturbance term in the reduced form of the model exist . Then , as proved in Dunsmuir (1979) , the quasi–MLE is consistent though its asymptotic variance is increased . Hence , we conclude that the M–estimators are more robust than the MLE .

4.5.3 Non–normal contaminating component

Additive outliers in a series can be generated from other mixture distributions beside the contaminated normal distribution . We shall investigate the performance of the M–estimators in comparison with MLEs when the contaminating component in the mixture distribution is non–normal . Let us first fix the hyperparameter values at 2 for both σ_ϵ^2 and σ_η^2 , i.e. we are considering the case where q = 1 . Suppose that the disturbance term in the measurement equation is given by

$$p(\epsilon_t) = (1-\gamma)\, N(0,\sigma_\epsilon^2) + \gamma\, \omega_t \qquad (4.5.12)$$

where the contaminating component ω_t follows a t–distribution with 2 degrees of freedom in case (a) and a Cauchy distribution in case (b) . The percentage of outliers is again fixed at 5% for convenience . One hundred random series each of length a

hundred are generated from the local level model in each case . M–estimates as well as ML estimates are then obtained for each replication . Sampling distributions for M–estimators are found in figures 4.20 to 4.21 while those of the MLE are given in figures 4.22 to 4.23 . Tables 4.6 and 4.7 contain the summary statistics derived from the simulation results.

TABLE 4.6 SUMMARY STATISTICS FOR M–ESTIMATORS
(non–normal contamination, γ=5%)

Distribution of ω_t		Bias	MSE
(i) Estimate of ln σ_ϵ (0.3466)			
(a)	t_2	–0.0674	0.0708
(b)	Cauchy	–0.1047	0.1315
(ii) Estimate of ln $\sqrt{q}$ (0.0)			
(a)	t_2	0.0026	0.1943
(b)	Cauchy	0.0607	0.3603

TABLE 4.7 SUMMARY STATISTICS FOR MLE
(non–normal contamination, γ=5%)

Distribution of ω_t		Bias	MSE	Bias*	MSE*	M
(i) Estimate of ln σ_ϵ (0.3466)						
(a)	t_2	–0.1816	3.2802	0.0074	0.0154	1
(b)	Cauchy	–0.0811	3.1219	0.0694	0.1003	3
(ii) Estimate of ln $\sqrt{q}$ (0.0)						
(a)	t_2	0.1590	3.5522	–0.0240	0.0774	1
(b)	Cauchy	0.0830	4.6097	–0.1330	0.2083	3

Table 4.7 indicates that the non–normal contaminating component causes the ML estimation procedure to breakdown . The number of breakdown cases is more for case (b) where the outliers are more extreme being generated from the Cauchy distribution . Such breakdowns do not occur with the M–estimation technique . In fact , the summary statistics for the M–estimators reveal that the hyperparameters are still reasonably well estimated . Hence , we say that the MLE are unable to cope with extreme outliers in terms of scale estimation and that the M–estimators are more robust than the MLE even when the outliers are not generated from the contaminated normal distribution .

4.5.4 Median estimators of scale

Suppose the series under consideration has a high percentage of contamination, i.e. with a big value for γ , running the M–estimation procedure described in the

previous section will result in big standard deviations for the M–estimators . To deal with heavily contaminated series , a smaller bound for truncation b , is used so that the derivative of more observations are modified . In the limiting case , b is chosen to be zero in order that the derivative of all the observations are adjusted . In fact , only their direction is taken into account during the computation . Hence , the Huber function in (4.4.14) is now replaced by

$$h(\underset{\sim}{z},t) = \frac{\underset{\sim}{z}}{\|\underset{\sim}{z}\|} \tag{4.5.13}$$

where the elements of $\underset{\sim}{z}$ remain the same as those in section 4.4.2. We call this new estimator the median estimator of scale since it corresponds to the median in the location estimation problem .

As before , the psi–functions constituting the vector $\underset{\sim}{z}$ have to be shifted to attain Fisher consistency . These calibration terms are different for different truncation bound values and they can be found by using the procedure described in section 4.5.1 . Results for median estimators are tabulated in Table 4.8 and the functional forms of c_1 and c_2 turn out to be similar to those of the M–estimators , see figures 4.24 and 4.25 .

TABLE 4.8 CALIBRATION RESULTS FOR MEDIAN ESTIMATOR

$\sqrt{q}$	σ_ϵ	c_1	c_2
0.1	1.00	1.2963	0.7137
	1.75	1.2976	0.6827
	2.25	1.3034	0.7088
0.3162	1.00	1.2332	0.6427
	1.75	1.2358	0.6738
	2.25	1.2446	0.6576
0.7071	1.00	1.1874	0.6141
	1.75	1.1756	0.6089
	2.25	1.1677	0.6052
1.0	1.00	1.1213	0.5707
	1.75	1.1450	0.5788
	2.25	1.1105	0.5659
2.0	1.00	1.0930	0.5490
	1.75	1.0876	0.5453
	2.25	1.0901	0.5469
3.1623	1.00	1.0883	0.5459
	1.75	1.0920	0.5459
	2.25	1.0947	0.5494
10.0	1.00	1.0817	0.5407
	1.75	1.0845	0.5426
	2.25	1.0765	0.5386

These calibration results indicate that c_1 and c_2 can be expressed in terms of the parameters σ_ϵ and $\sqrt{q}$ in the following relations

$$c_1 = 1.020 + 0.101/(\sqrt{q})^{0.3} \qquad (4.5.14a)$$

$$c_2 = 0.497 + 0.0752/(\sqrt{q})^{0.3} \qquad (4.5.14b)$$

Comparing these equations with those in (4.5.3) , we see that the calibration terms for median estimators are hardly different from those for the M–estimators with b = 0.5 .

To investigate the performance of the median estimators , we repeat the simulations in section 4.5.2 first for the case of clean data . Table 4.9 below consists of the summary statistics for the estimators in the absence of contamination . Their sampling distributions are found in figures 4.26 to 4.28 .

TABLE 4.9 SUMMARY STATISTICS FOR MEDIAN–ESTIMATORS (clean data)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	–0.1201	0.0982
(b)	0.3466	–0.1657	0.1433
(c)	0.0	–0.2251	0.3039
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.2737	0.4822
(b)	0.0	–0.2089	0.3853
(c)	0.2027	–0.1622	0.5433

The bias and mean square errors for the median estimators are significantly bigger than the corresponding values for the MLEs in the estimation of both $\ln \sigma_\epsilon$

and ln $\sqrt{q}$. In fact , all the summary statistics here are bigger than those of the M–estimators , see Table 4.2 . On the comparison of their sampling distributions , we see that the median estimators have longer tails than the MLEs and the M–estimators . This is because there is a substantial loss in information when the median estimation technique only uses the direction of $\underset{\sim}{z}$, see equation (4.5.13) .

Next , we repeat the simulations in section 4.5.2 for the case of additive outlier contamination but with the percentage of contamination increased to 10% . The relation between k and q remains as

$$k = 16.2 + 7.6q \qquad (4.5.15)$$

Hence , the value of k for the three cases are 20 , 25 and 30 . Simulation results for both median estimators and MLEs are summarized in Tables 4.10 and 4.11 below . Their sampling distributions are found in figures 4.29 to 4.34 .

TABLE 4.10 SUMMARY STATISTICS FOR MEDIAN–ESTIMATORS (γ=10%)

Estimator of		Bias	MSE
(i)	ln σ_ϵ		
(a)	0.3466	0.1159	0.0685
(b)	0.3466	0.1034	0.1413
(c)	0.0	0.1367	0.1032
(ii)	ln $\sqrt{q}$		
(a)	–0.3466	–0.1534	0.3902
(b)	0.0	–0.1083	0.4713
(c)	0.2027	–0.0908	0.2564

TABLE 4.11 SUMMARY STATISTICS FOR MLE (γ=10%)

Estimator of		Bias	MSE
(i)	$\ln \sigma_\epsilon$		
(a)	0.3466	0.4879	0.2830
(b)	0.3466	0.5572	0.3672
(c)	0.0	0.6203	0.4513
(ii)	$\ln \sqrt{q}$		
(a)	–0.3466	–0.5952	0.5076
(b)	0.0	–0.6418	0.5639
(c)	0.2027	–0.7011	0.6435

Here , we see the median estimators having a definite edge over the MLEs in terms of their summary statistics . Both the bias and mean square errors for the MLEs are significantly bigger than the corresponding values for the median estimators .

We next repeat the simulations in section 4.5.3 to compare the performance of median estimators with that of MLEs when the contaminating component is non–normal . However , the percentage of contamination is now held at 10% . Simulation results are given in Tables 4.12 and 4.13 and the sampling distributions are found in figures 4.35 to 4.38 .

TABLE 4.12 SUMMARY STATISTICS FOR MEDIAN ESTIMATORS
(non–normal contamination, γ=10%)

Distribution of ω_t		Bias	MSE
(i) Estimate of ln σ_ϵ (0.3466)			
(a)	t_2	–0.1297	0.0918
(b)	Cauchy	–0.1566	0.1848
(ii) Estimate of ln $\sqrt{q}$ (0.0)			
(a)	t_2	0.1111	0.2197
(b)	Cauchy	0.1505	0.4601

TABLE 4.13 SUMMARY STATISTICS FOR MLE
(non–normal contamination, γ=10%)

Distribution of ω_t		Bias	MSE	Bias*	MSE*	M
(i) Estimate of ln σ_ϵ (0.3466)						
(a)	t_2	–0.1416	2.6025	–0.0096	0.0310	1
(b)	Cauchy	0.1436	9.0266	0.2254	0.3270	4
(ii) Estimate of ln $\sqrt{q}$ (0.0)						
(a)	t_2	0.1220	2.8642	–0.0220	0.2273	1
(b)	Cauchy	0.0010	16.4511	–0.3510	0.5193	4

Again , we observe that both the MLE break down with non–normal contamination , see Table 4.7 . On the other hand , the median estimators do not break down even for the case where there is 10% contamination modelled by the

Cauchy distribution . Hence , we conclude from the simulation results that the median estimator is more robust towards outliers when compared with the MLE . It is recommended for use with contaminated data especially when the percentage of contamination is suspected to be high .

4.6 Redescending estimators

4.6.1 The estimation procedure

Another robust estimator of scale is the redescending estimator which is designed to deal with extreme outliers . Consider a univariate M–estimator of scale , for example , the one given in (4.2.7) . A redescending estimator is similar to this Huber curve in that the psi–function remains unmodified if it is small . However , if the influence of an observation is too big , the psi–function is no longer replaced by a constant . Instead , it is downweighted by a function which goes to zero . Thus , when the influence of an observation exceeds a bound r , its psi–function is given zero weight . Andrew's curve AC(x) uses a sine function to downweight the influence of outliers and it is given by

$$AC(x) = \frac{r}{\pi} \sin\left(\frac{\pi}{r} x\right) \qquad 0 \leq x \leq r \tag{4.6.1}$$

where r acts as a tuning constant which determines the width of the window . We note that the sine function is like the identity function when its argument is very small .Hence , for very small values of x ,

$$AC(x) = x \tag{4.6.2}$$

Redescending estimators for the parameters σ_ϵ and $\sqrt{q}$ are then defined as the roots of the following equation

$$\sum_{t=1}^{T} R(\underset{\sim}{z},t) = 0 \tag{4.6.3a}$$

where $R(\underset{\sim}{z},t)$ corresponds to the Huber function $h(\underset{\sim}{z},t)$ in (4.4.14) and is given by

$$R(\underset{\sim}{z},t) = \begin{cases} \dfrac{\underset{\sim}{z}}{\|\underset{\sim}{z}\|} \dfrac{r}{\pi} \sin\left[\dfrac{\pi}{r}\|\underset{\sim}{z}\|\right] & \|\underset{\sim}{z}\| \leq r \\ \underset{\sim}{0} & \|\underset{\sim}{z}\| > r \end{cases} \tag{4.6.3b}$$

with $\underset{\sim}{0} = (0\ 0)'$

and $\underset{\sim}{z}$ comprises the two psi–functions $\Psi^*(\sigma_\epsilon,t)$ and $\Psi^*(\sqrt{q},t)$ as in (4.4.14) . When the magnitude of $\underset{\sim}{z}$ is very small , the function $R(\underset{\sim}{z},t)$ acts as an identity function and the psi–functions in $\underset{\sim}{z}$, are not modified . As the magnitude of $\underset{\sim}{z}$ increases , both psi–functions have a weight which is less than one and it declines to zero . All observations whose influence $\|\underset{\sim}{z}\|$ exceeds the cut–off r are given zero weight , i.e. , they are not taken into account in the sum (4.6.3a) . In this way , extreme outliers are discarded during scale estimation and the resulting estimators are more robust . However , such redescending estimation procedure often produce multiple roots .

4.6.2 Performance of redescending estimators

The calibration process described in section 4.5.1 is repeated here except for the replacement of the Huber function $h(\underset{\sim}{z},t)$ by the redescending function $R(\underset{\sim}{z},t)$ and r is fixed at five . The values of c_1 and c_2 for different sets of parameters values are tabulated below

TABLE 4.14 CALIBRATION RESULTS FOR REDESCENDING ESTIMATORS

$\sqrt{q}$	σ_ϵ	c_1	c_2
0.1	1.00	1.4566	0.9113
	1.75	1.4565	0.8828
	2.25	1.4650	0.9172
0.3162	1.00	1.3386	0.8169
	1.75	1.3394	0.8567
	2.25	1.3436	0.8347
0.7071	1.00	1.2000	0.7035
	1.75	1.1943	0.7040
	2.25	1.1815	0.6966
1.0	1.00	1.0698	0.5650
	1.75	1.0820	0.5687
	2.25	1.0681	0.5654
2.0	1.00	1.0407	0.5266
	1.75	1.0578	0.5249
	2.25	1.0411	0.5250
3.1623	1.00	1.0340	0.5188
	1.75	1.0405	0.5212
	2.25	1.0413	0.5226
10.0	1.00	1.0390	0.5195
	1.75	1.0370	0.5185
	2.25	1.0349	0.5177

A plot of c_1 and c_2 against the parameters σ_ϵ and $\sqrt{q}$ are found in figures 4.39 and 4.40 . We see from figure 4.39a and 4.40a that the calibration terms are independent

of the the parameter σ_ϵ . If we shift the curves in figures 4.39b and 4.40b down so that they tend to the origin , non–linear optimisation gives the following relations of c_1 and c_2 in terms of $\sqrt{q}$.

$$c_1 = 1.0 + 1./(1.8 + 10.3\sqrt{q}^{0.94}) \qquad R^2 = 99.4\% \qquad (4.6.4a)$$

$$c_2 = 0.5 + 1./(2.3 + 12.2\sqrt{q}^{1.3}) \qquad R^2 = 99.5\% \qquad (4.6.4b)$$

Like the functions in (4.5.3) , the above relations are a good fit to the sampling results as indicated by the high R^2 values . These relations are also confirmed by the fact that the actual values of c_1 and c_2 are almost identical to their fitted values , see figures 4.39c and 4.40c . Hence , they are used in the simulations below .

To check the performance of redescending estimators , the simulations in section 4.5.2 and 4.5.3 are repeated for this new estimation procedure . The series which have been generated previously for the three cases are used here . For each series , we have to obtain starting values for the hyperparameters as in Chapter 3 . Like the M–estimators , we have to use different starting values for cases of non–convergence , see pg 154 . With the tuning constant r fixed at five and replacing c_1 and c_2 in (4.6.3) by the functions in (4.6.4) , we have the following results :

TABLE 4.15 SUMMARY STATISTICS FOR REDESCENDING ESTIMATORS
(clean data)

Estimator of		Bias	MSE	Bias*	MSE*	M
(i)	$\ln \sigma_\epsilon$					
(a)	0.3466	–0.2153	0.4989	–0.1267	0.1876	1
(b)	0.3466	–0.1993	0.2840			
(c)	0.0	–0.2907	0.7684	–0.1697	0.2335	3
(ii)	$\ln \sqrt{q}$					
(a)	–0.3466	–0.7640	1.8833	–0.6080	1.3711	1
(b)	0.0	–0.4490	1.3337			
(c)	0.2027	0.0303	1.8478	–0.1647	0.9630	2

The sampling distributions of the above estimators are given in figures 4.41 to 4.43 . From the above table , we see that redescending estimators sometimes breakdown even when the data is clean . This is due to multiple solutions associated with redescending estimators . A comparison of these results with those in Table 4.3 show that the redescending estimators have bigger bias than the MLEs especially for the estimator of $\ln\sqrt{q}$. The figures reveal that the sampling distribution of redescending estimators have rather long tails . Thus , the mean square errors for the MLEs are significantly smaller than those for the redescending estimators with clean data .

TABLE 4.16 SUMMARY STATISTICS FOR REDESCENDING ESTIMATORS (γ=5%)

Estimator of		Bias	MSE	Bias*	MSE*	M
(i)	$\ln \sigma_\epsilon$					
(a)	0.3466	–0.1758	0.7031	0.0075	0.2253	2
(b)	0.3466	–0.1519	0.5779	0.0451	0.0849	1
(c)	0.0	–0.0878	0.5632	0.0518	0.0978	2
(ii)	$\ln \sqrt{q}$					
(a)	–0.3466	–0.4924	2.0381	–0.3814	0.8213	2
(b)	0.0	–0.3490	1.5001	–0.2680	0.4993	1
(c)	0.2027	–0.3537	1.4964	–0.2987	0.4480	2

The sampling distributions of the above estimators are given in figures 4.44 to 4.46 . We observe that there are either one or two breakdowns in each of the cases above . By looking at the resistant set of summary statistics , we see that $\ln \sqrt{q}$ is poorly estimated because of its big MSE* . The sampling distributions for the estimator of this parameter are generally long–tailed . In comparison with the MLEs , we say that the redescending estimators are more robust in that there is no significant shift in their sampling distributions . Thus , the biases of redescending estimators tend to be smaller than those of the MLEs . However , the sampling distributions of the MLEs are much tighter than those of the redescending estimators .

TABLE 4.17 SUMMARY STATISTICS FOR REDESCENDING ESTIMATORS
(non–normal contamination, γ=5%)

Distribution of ω_t		Bias	MSE	Bias*	MSE*	M
(i) Estimate of ln σ_ϵ (0.3466)						
(a)	t_2	–0.2337	0.6586	–0.0322	0.1133	2
(b)	Cauchy	–0.1990	0.4568	–0.0703	0.0895	1
(ii) Estimate of ln $\sqrt{q}$ (0.0)						
(a)	t_2	–0.2610	1.4795	–0.1130	0.3194	1
(b)	Cauchy	–0.3300	1.4966	–0.2080	0.6526	1

The sampling distributions of the above estimators are given in figures 4.47 and 4.48 . The redescending estimators behave in a similar way regardless of the distribution of the contaminating component . There are one or two breakdowns in the cases above and $\ln\sigma_\epsilon$ is better estimated than $\ln\sqrt{q}$. If we compare the above results with those of case (b) in Table 4.16 , we see that the redescending estimators are quite insensitive towards the type of contamination . When comparing with the MLEs , we still observe tighter sampling distributions for the MLEs . However , there are fewer cases of breakdown for the redescending estimators when the outliers are generated from the Cauchy distribution .

Since the M–estimators do not breakdown so easily , they are preferred to redescending estimators in obtaining robust estimates of scale .

4.7 Other Estimators

A structural time series model can be expressed as an autoregressive integrated moving average (ARIMA) model . In particular , the local level model corresponds to a ARIMA(0,1,1) model . That is , it becomes a MA(1) or a first order moving average model after taking first differences on the observations . Hence , from the y_t in (4.1.1) we have

$$\begin{aligned} \Delta y_t &= \eta_t + \epsilon_t - \epsilon_{t-1} \\ &= \xi_t - \delta\xi_{t-1} \end{aligned} \tag{4.7.1}$$

It follows that there is a functional relation between the hyperparameters of a structural model and the parameters in a ARIMA model . The following relations can be derived from equation (4.7.1)

$$\sigma_\xi^2 = \frac{\sigma_\epsilon^2}{\delta} \tag{4.7.2a}$$

$$\delta = \frac{(2+q) \pm \sqrt{q^2+4q}}{2} \tag{4.7.2b}$$

This implies that the estimation of σ_ϵ^2 and q in the local level model is linked to the estimation of σ_ξ^2 and δ in the MA(1) model .

Various robust estimators have been proposed for the estimation of parameters in ARIMA models , see Martin and Yohai (1985) . We shall discuss these estimators

in the context of a MA(1) model which is given by

$$y_t = u_t - \theta u_{t-1} \tag{4.7.3}$$

The parameter of interest here is θ while σ^2 , the variance of the disturbance u_t , is often regarded as a nuisance parameter and is estimated from the residuals $\hat{u}_t(\theta)$ in an iterative process . The least squares estimate of θ is obtained by minimising

$$\sum_{t=1}^{T} \hat{u}_t^2(\theta) \tag{4.7.4a}$$

where

$$\hat{u}_t(\theta) = \frac{1}{1-\theta L} y_t$$

$$= \sum_{i=0}^{\infty} t_i L^i y_t \tag{4.7.4b}$$

The lag operator is denoted by L . M–estimates are then , defined as the solution to

$$\min_{\theta} \sum_{t=1}^{T} \rho(\hat{u}_t(\theta)) \tag{4.7.5a}$$

or

$$\sum_{t=1}^{T} \psi\left[\frac{\hat{u}_t(\theta)}{\sigma}\right] b_t(\theta) = 0 \tag{4.7.5b}$$

where

$$b_t(\theta) = \frac{1}{1-\theta L} \hat{u}_t(\theta) \tag{4.7.5c}$$

Note that ψ is the derivative of the function ρ and can be the Huber function defined in Chapter 2 , equation (2.6.3) . Both the least squares and M–estimators are not robust since their sums in (4.7.4) and (4.7.5) are not bounded . To accomodate the

outliers , three different robust estimators are proposed . They are

(a) General M–estimates

GM–estimates is first introduced for the robust estimation of regression parameters and takes the form of weighted least squares . It is the root of the following equation

$$\sum_{t=1}^{T} \omega\left[\frac{\hat{u}_t(\theta)}{\sigma}, \frac{b_{t-1}^2}{\hat{C}}\right] \hat{u}_t(\theta) b_{t-1}(\theta) = 0 \qquad (4.7.6a)$$

where $\hat{C}$ is an estimate of the variance of $b_{t-1}(\theta)$. The ω function usually assumes one of the following forms

(i) Mallows type

$$\omega(u,v) = \frac{\psi_1(u)\psi_2(v^{1/2})}{uv^{1/2}} \qquad (4.7.6b)$$

(ii) Hampel–Krasker–Welsch type

$$\omega(u,v) = \frac{\psi(uv^{1/2})}{uv^{1/2}} \qquad (4.7.6c)$$

(b) Residual Autocovariance (RA) estimates

Using the equation (4.7.4b) , the M–estimates in (4.7.5) can be rewritten as

$$\sum_{h=0}^{T-2} t_h(\theta)\hat{\gamma}_{h+1}(\theta) = 0 \qquad (4.7.7a)$$

where

$$\hat{\gamma}_i(\theta) = \sum_{t=1}^{T-i} \psi\left[\frac{\hat{u}_{t+i}(\theta)}{\sigma}\right] \hat{u}_t(\theta) \qquad (4.7.7b)$$

which is an estimate of the residuals autocovariance at lag i multiplied by a factor . In RA estimates , this is replaced by the following robust estimate

$$\hat{\gamma}_i(\theta) = \sum_{t=1}^{T-i} \eta\left[\frac{\hat{u}_{t+i}(\theta)}{\sigma} \cdot \frac{\hat{u}_t(\theta)}{\sigma}\right] \qquad (4.7.8)$$

Like the ω function in GM–estimates , the η function here can take on the Mallows type in (4.7.9a) or the Hampel type in (4.7.9b) below .

$$\eta(u,v) = \psi_1(u)\psi_2(v) \qquad (4.7.9a)$$

$$\eta(u,v) = \psi(uv) \qquad (4.7.9b)$$

Both the GM and RA estimators are mainly used to estimate parameters in a stationary ARMA model . We should not try to estimate the hyperparameters of the local level model through these estimators using the relations in (4.7.2) . This is because differencing of the observations in (4.7.1) will result in a proliferation of outliers .

(c) Approximate Maximum Likelihood estimates (AMLE)

This type of estimator is more appropriate for sequential estimation . The ARIMA model is first put in state space form and then a filter is used . As seen in

the previous chapters , eg. in (2.2.7) , the likelihood of a time series can be decomposed into its conditional densities . Thus , we can write

$$\ln L(Y_T) = \sum_{t=1}^{T} \ln g_t(v_t) \tag{4.7.10a}$$

where v_t is the prediction error while g_t is a convolution of $p(x_t/Y_{t-1})$ and $p(\epsilon_t)$. Under Masreliez assumption and letting the disturbance ϵ_t take on a long tailed distribution , the following approximation is carried out .

$$g_t(v_t) \approx \frac{1}{\sqrt{f_t}} g\left[\frac{v_t}{\sqrt{f_t}}\right] \tag{4.7.10b}$$

where f_t is the scale of the prediction errors . Thus , to obtain maximum likelihood estimates we minimise

$$\sum_{t=1}^{T} \ln \sqrt{f_t} - \sum_{t=1}^{T} \ln g\left[\frac{v_t}{\sqrt{f_t}}\right] \tag{4.7.10c}$$

AMLE are obtained by minimising the following robustified objective function

$$\sum_{t=1}^{T} \ln \sqrt{f_t} - \sum_{t=1}^{T} \rho\left[\frac{v_t}{\sqrt{f_t}}\right] \tag{4.7.11a}$$

Letting z be the ratio of v_t to $\sqrt{f_t}$ and c be the cut–off , one of the forms which ρ can take is

$$\rho(z) = \begin{cases} -cz - z^2/2 & z < -c \\ z^2/2 & \|z\| \leq c \\ cz - z^2/2 & z > c \end{cases} \qquad (4.7.11b)$$

In the control engineering literature , see Ljung (1978) , this procedure is known as the prediction–error method . Robustness is achieved by letting the ρ function "increase more slowly than quadratically" in v_t . However , this method is based on the assumption that f_t is independent of the parameters of interest . With reference to section 4.3.2 , we see that although v_t and f_t are independent of the parameter σ_ϵ , both $v_t$and f_t are functions of q , the exact relation is given in (4.4.3) and (4.4.6) . Hence , neither the approximation in (4.7.10b) nor the independence assumption is satisfied .

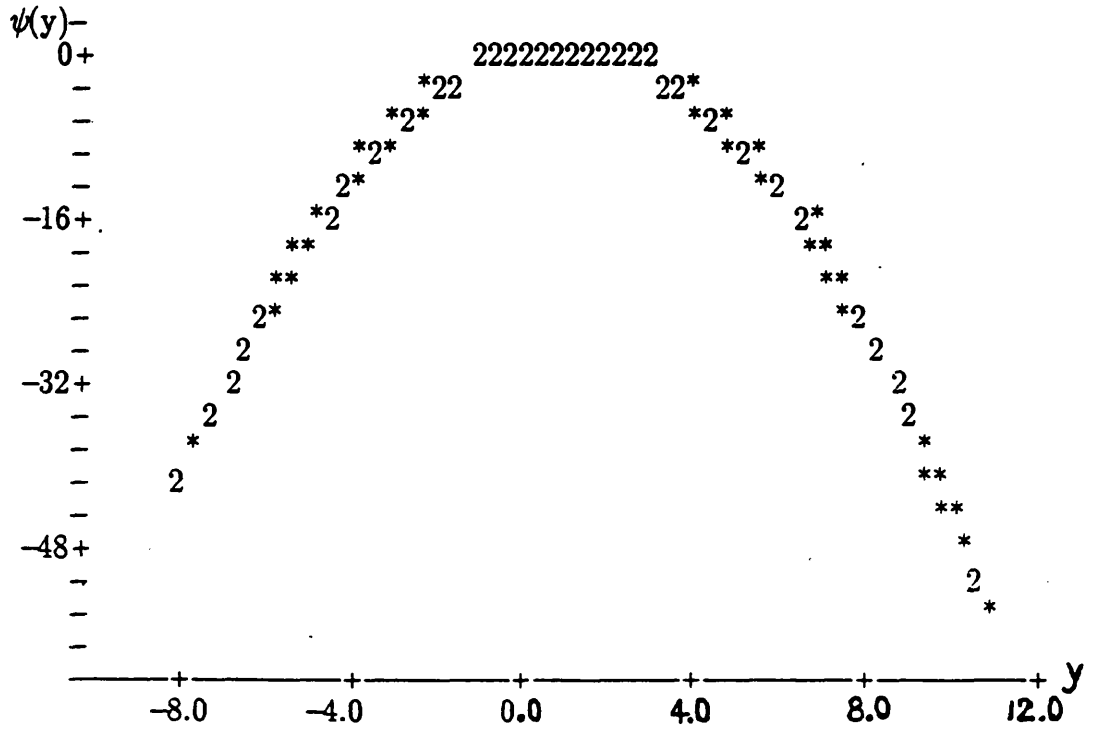


Figure 4.1 Psi–function for observations from N(1,2) distribution

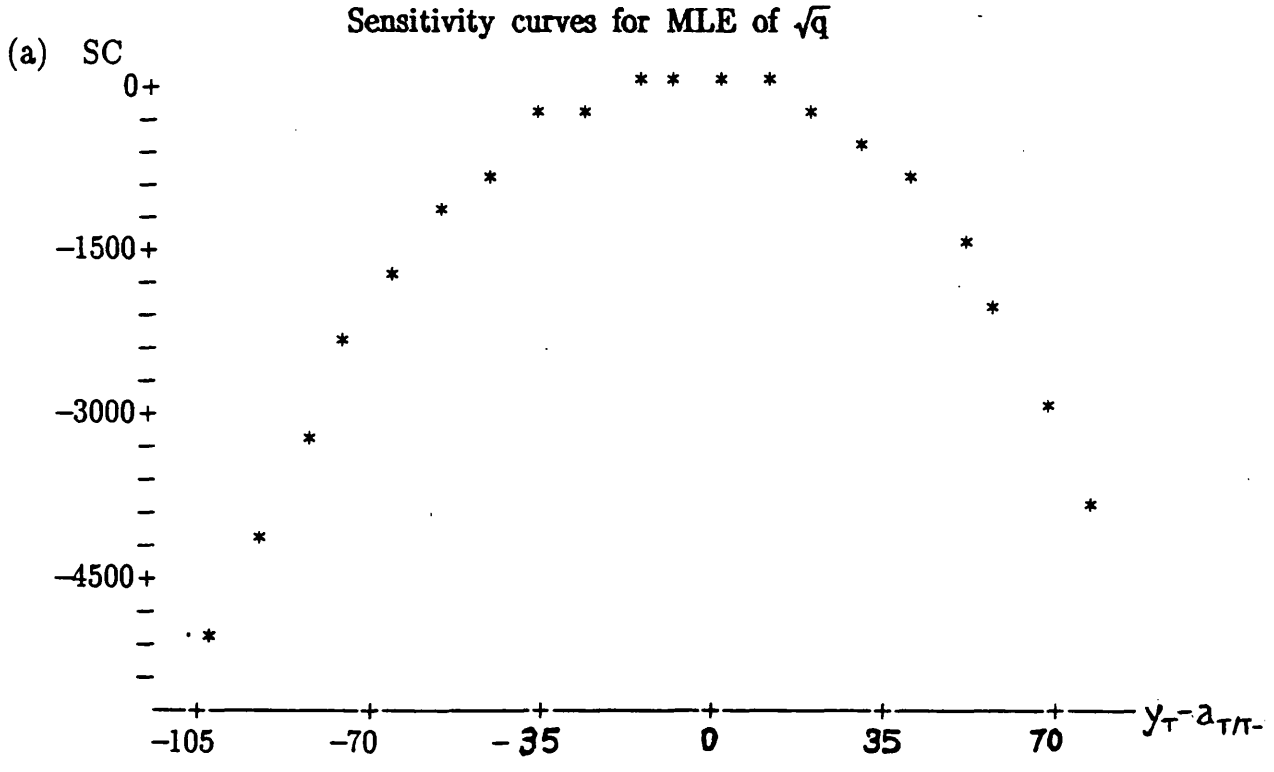


Figure 4.2a T = 300 and q = .5

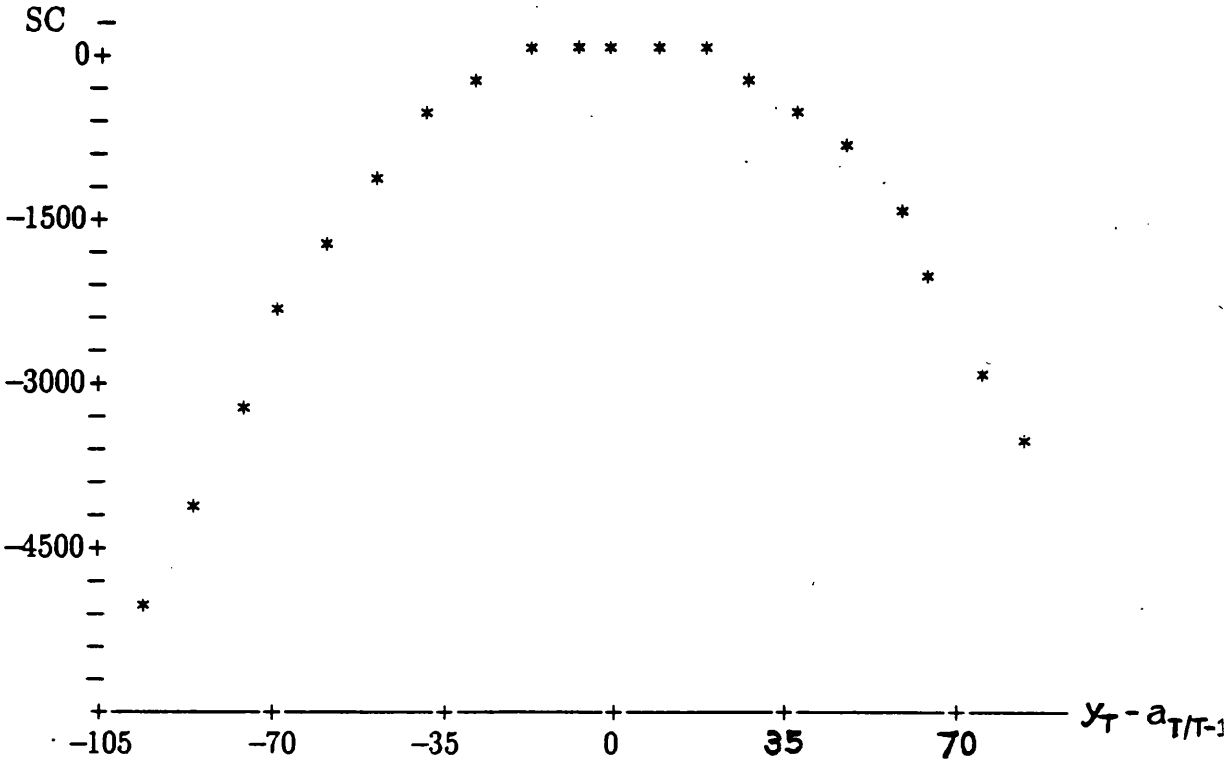


Figure 4.2b T = 400 and q = .5

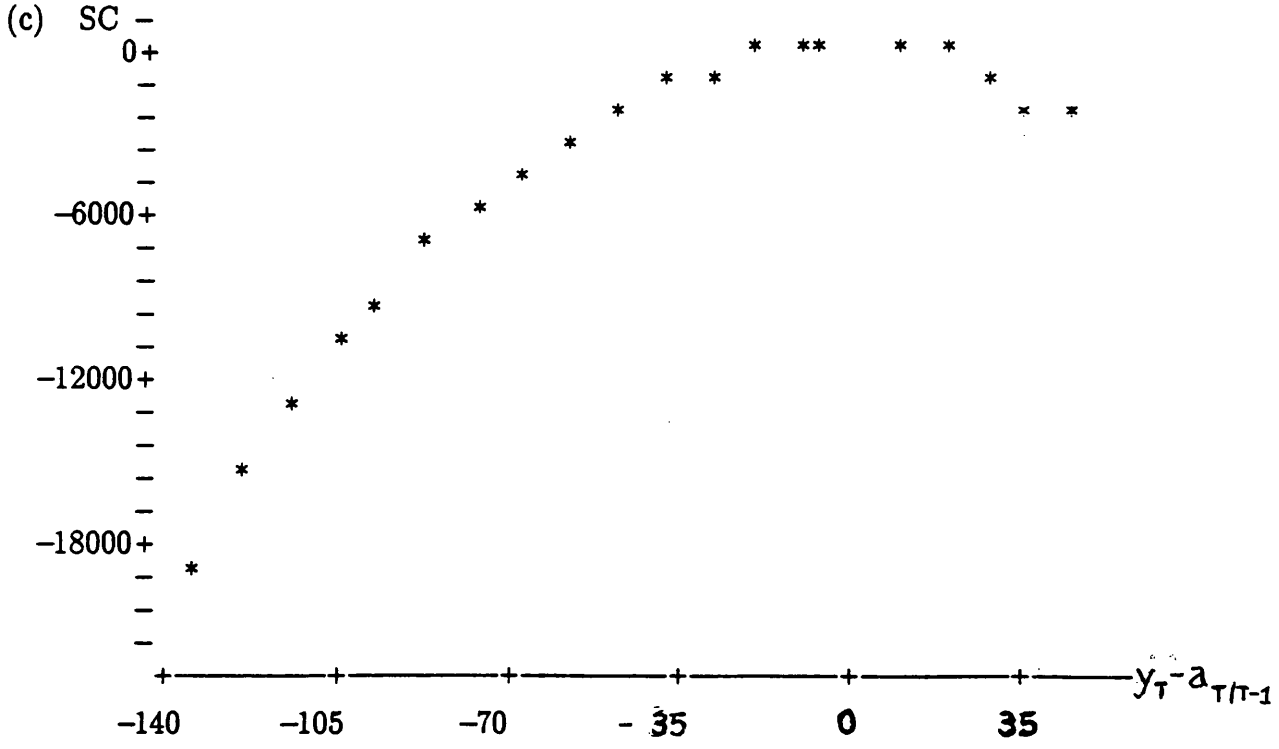


FIgure 4.2c T = 300 and q = 5.

Calibration term for $\psi(\sigma_\epsilon, t)$, c_1 of M–estimators

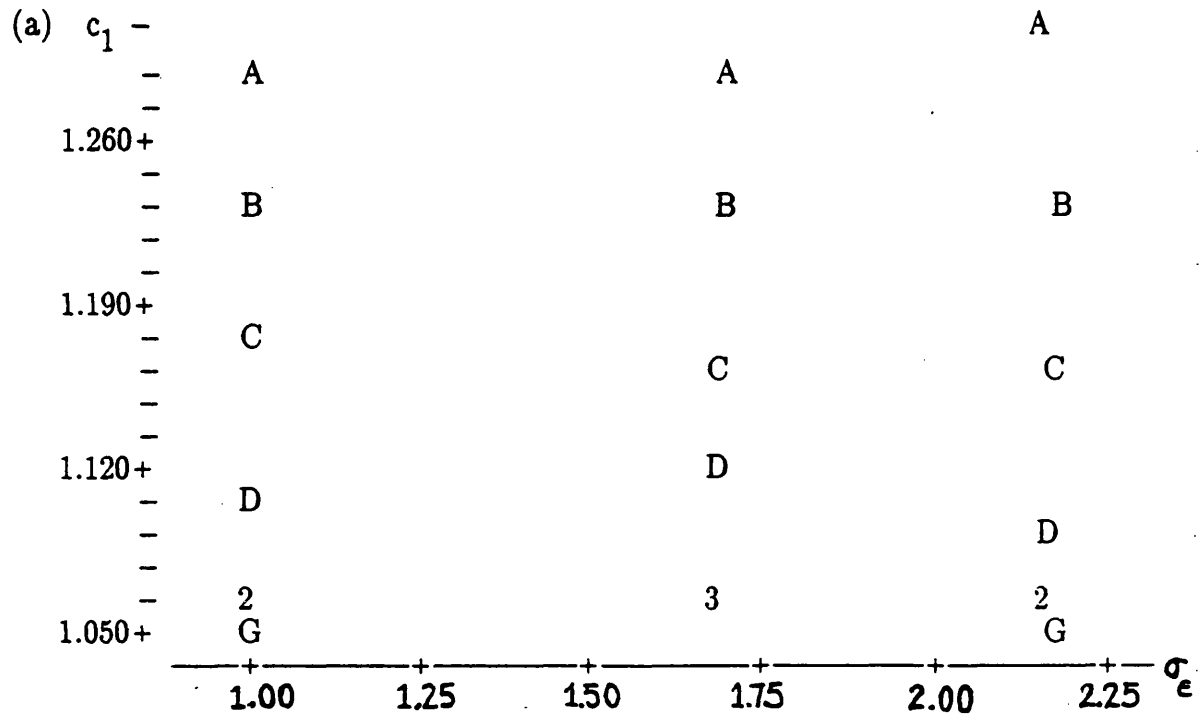


Figure 4.4a Plot c_1 against σ_ϵ

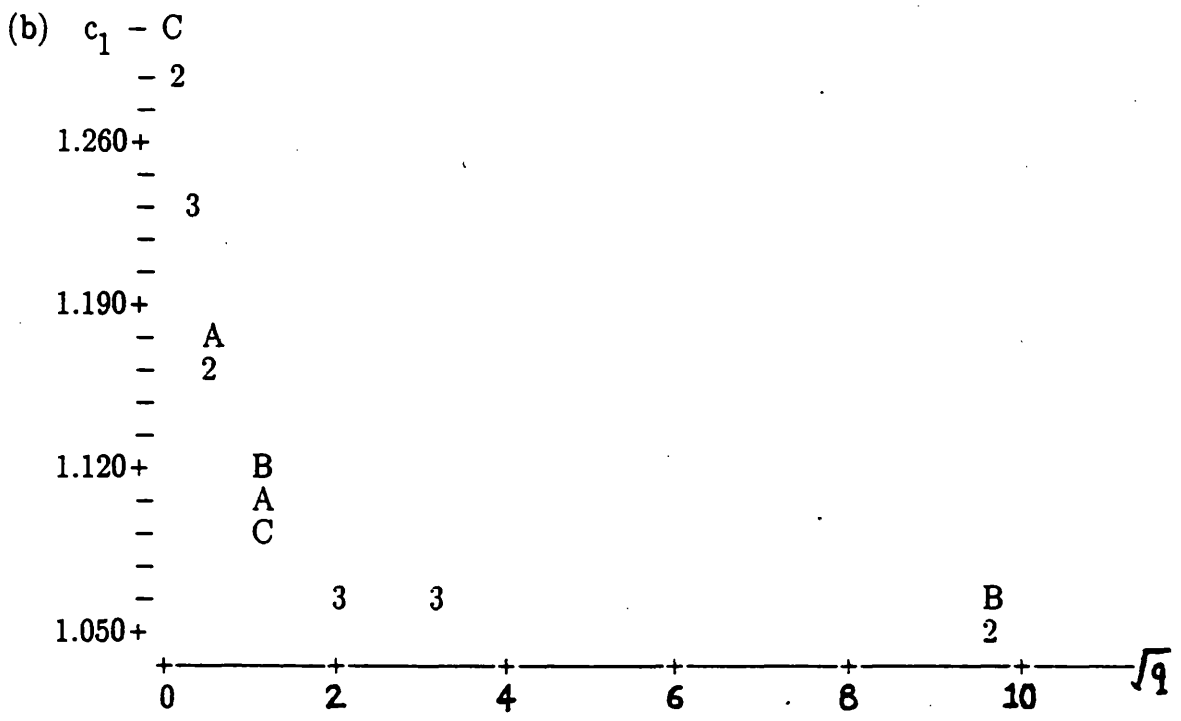


Figure 4.4b Plot of c_1 against $\sqrt{q}$

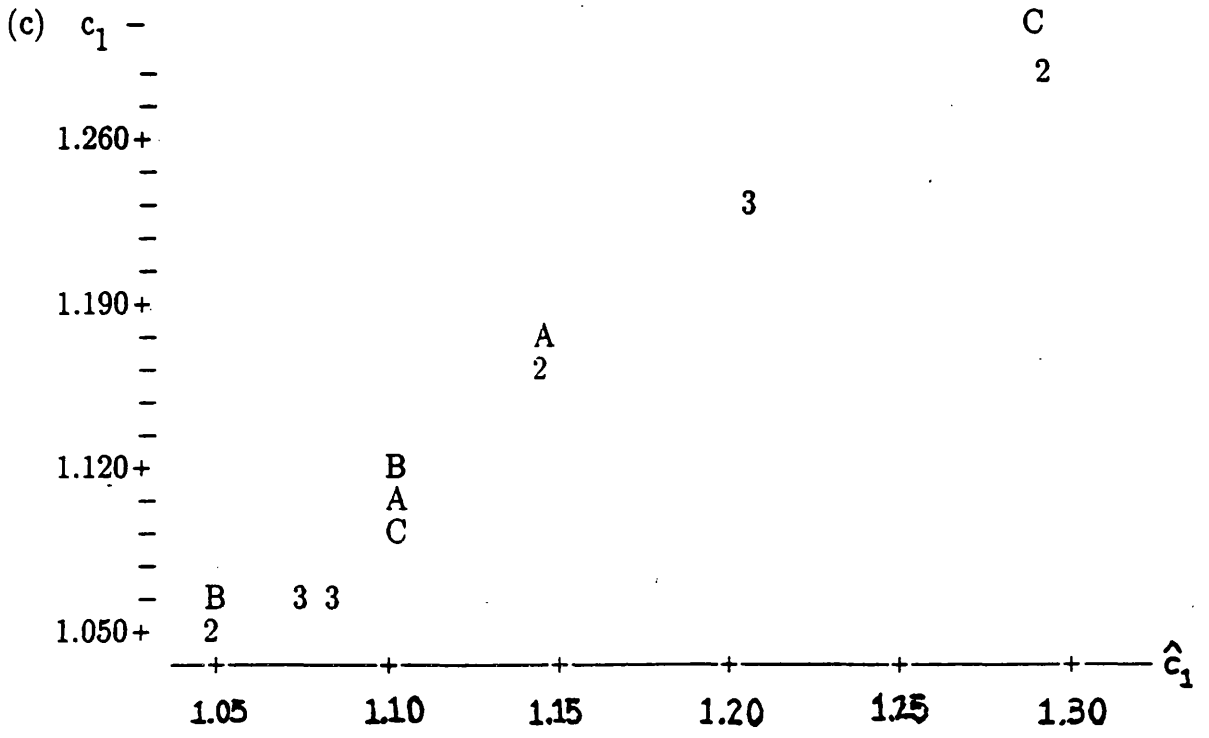


Figure 4.4c Plot c_1 against its fitted values

Calibration term for $\psi(\sqrt{q},t)$, c_2 of M–estimators

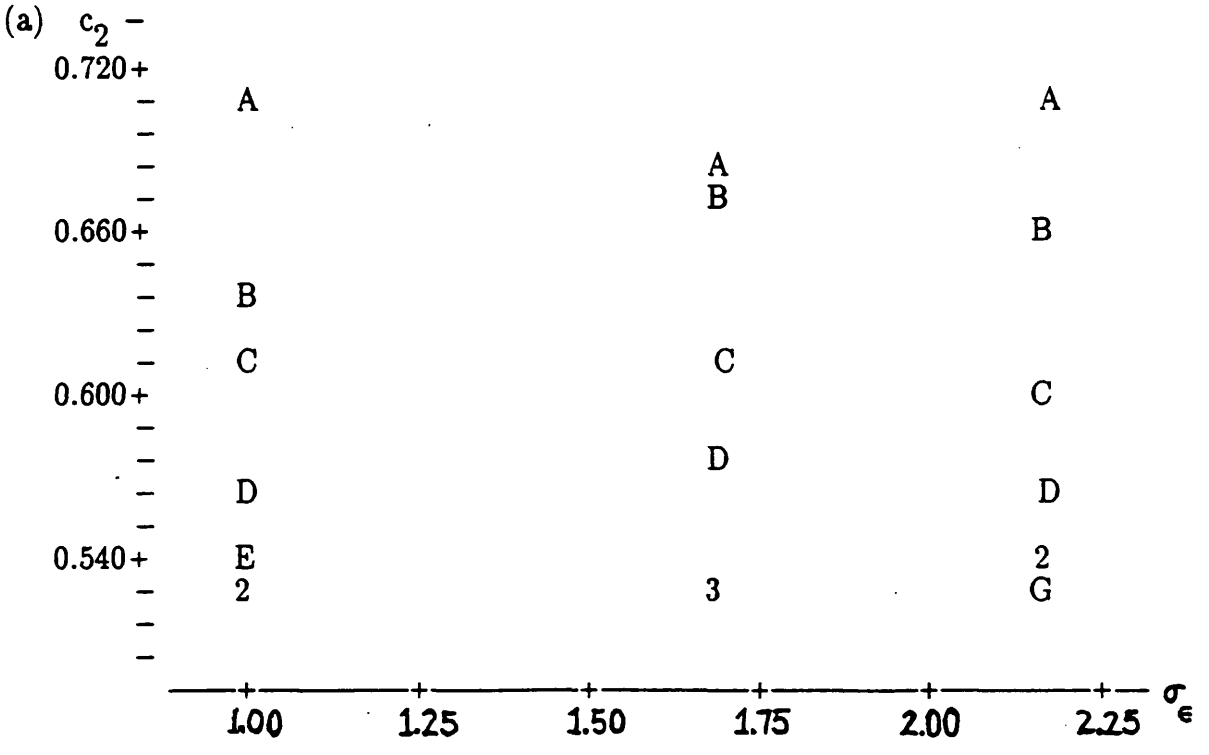


Figure 4.5a Plot of c_2 against σ_ϵ

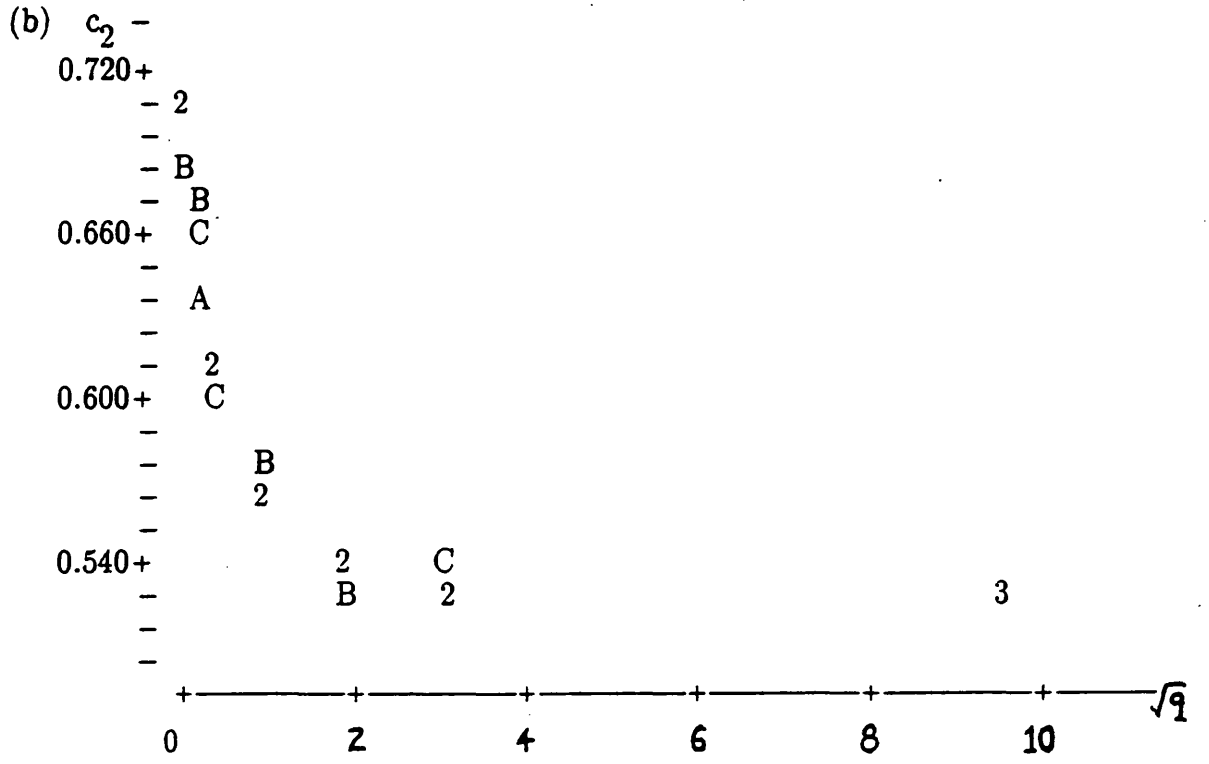


Figure 4.5b Plot of c_2 and $\sqrt{q}$

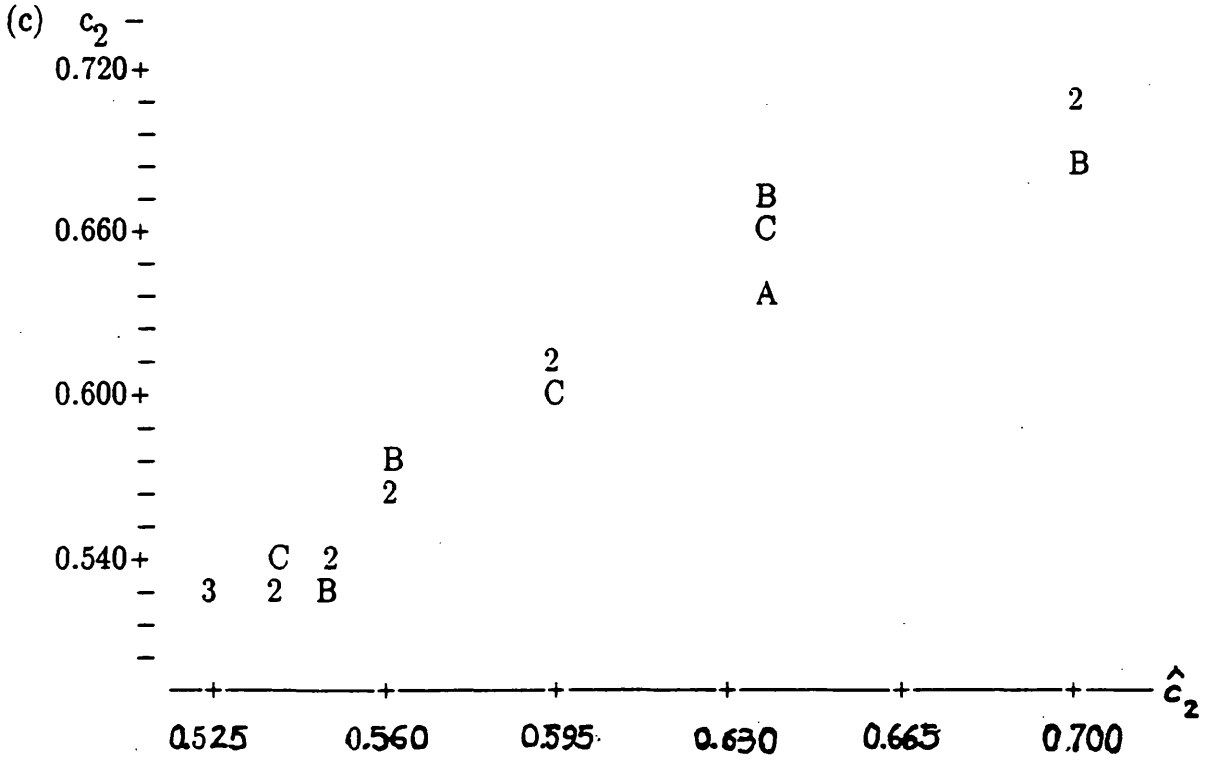


Figure 4.5c Plot of c_2 against its fitted values

Calibration terms c_1 and c_2 of M–estimators when T=200

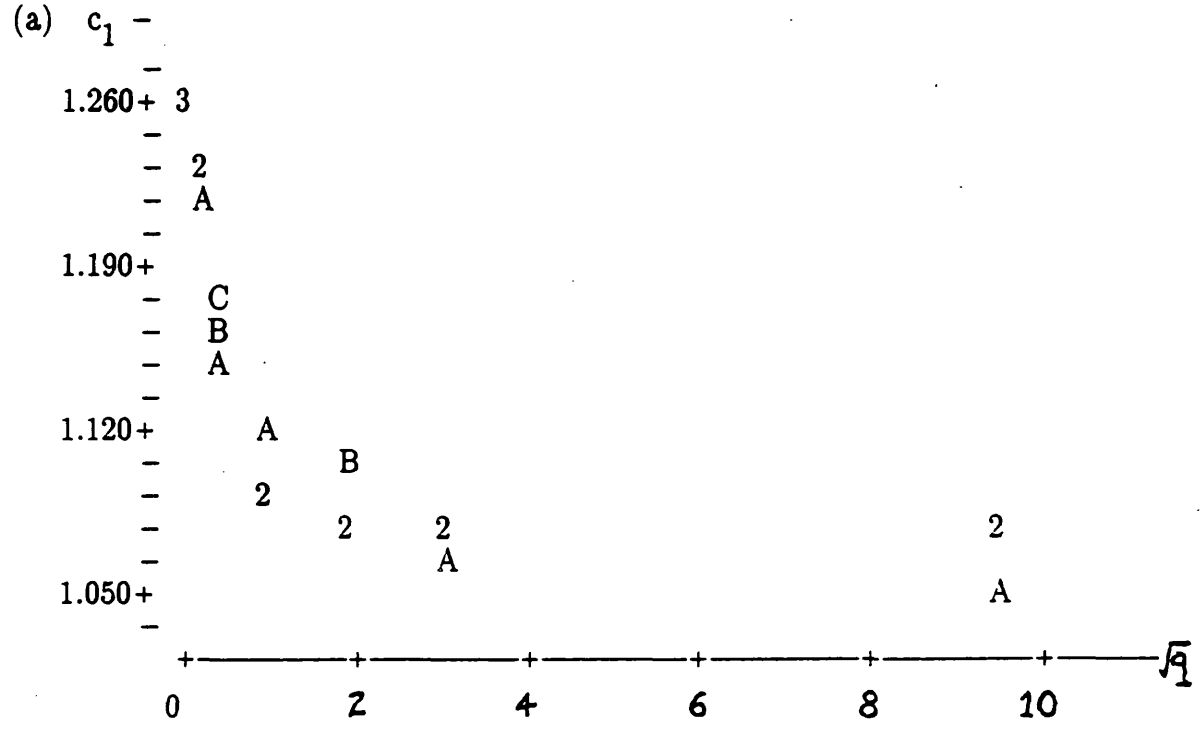


Figure 4.6a Plot of c_1 against $\sqrt{q}$

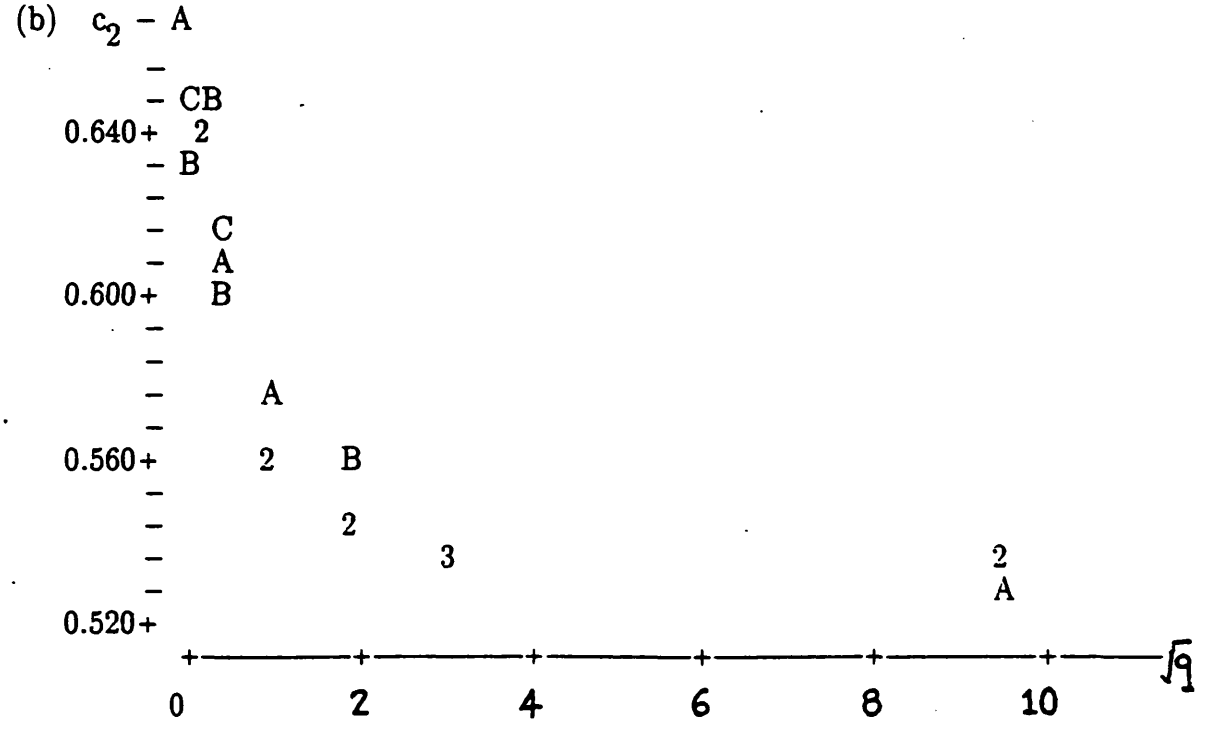


Figure 4.6b Plot of c_2 against $\sqrt{q}$

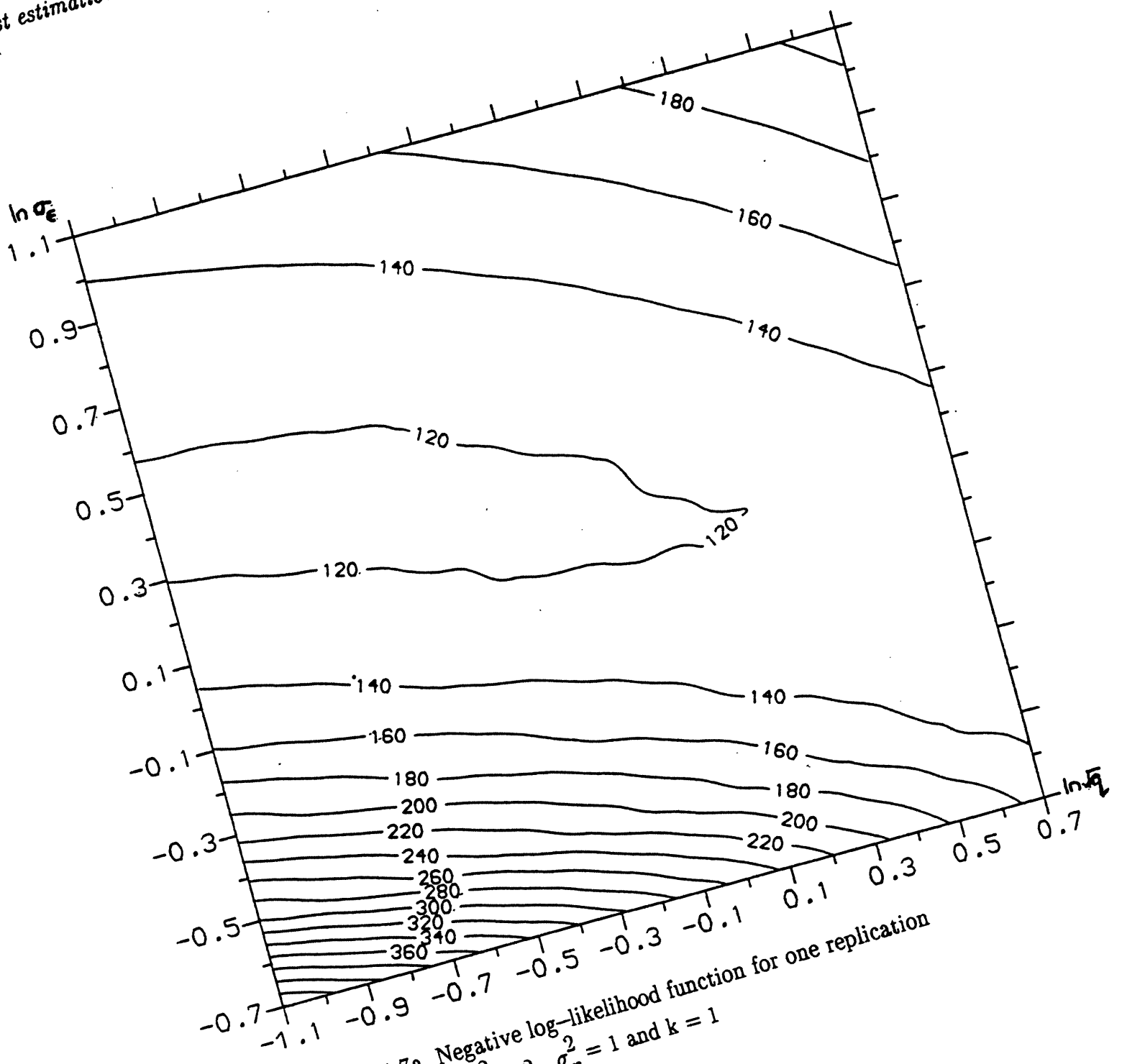


Figure 4.7a Negative log–likelihood function for one replication when $\sigma^2_\epsilon = 2$, $\sigma^2_\eta = 1$ and k = 1

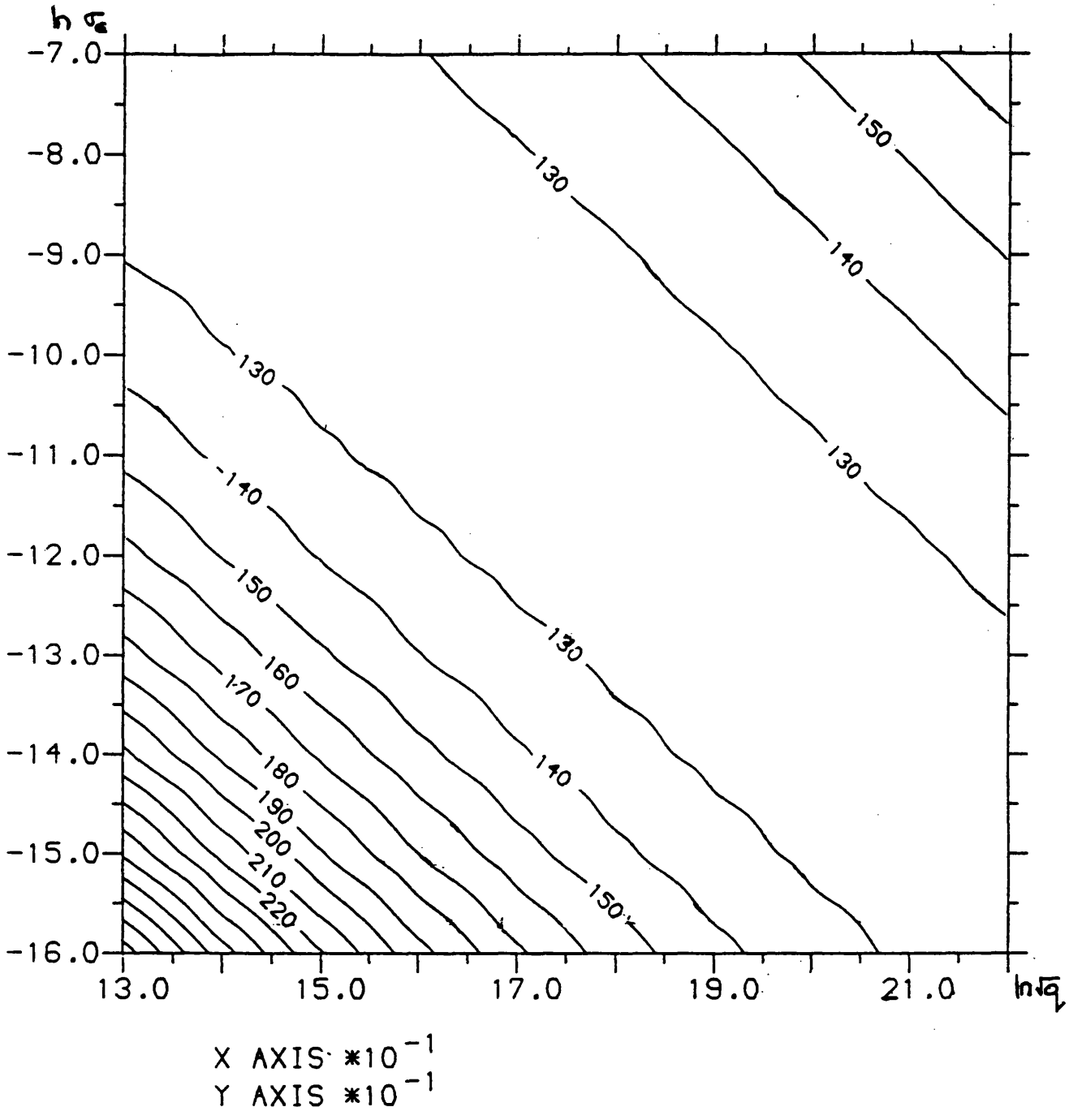


Figure 4.7b Lower right–hand corner of the objective function

Sampling distributions for M–estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for clean data

(a) Midpoint Count

–0.4	1	*
–0.2	4	****
0.0	13	**********
0.2	30	******************************
0.4	38	**************************************
0.6	13	**********
0.8	1	*

(b) Midpoint Count

–2.0	1	*
–1.5	3	***
–1.0	16	****************
–0.5	42	**
0.0	29	****************************
0.5	8	********
1.0	1	*

Figure 4.8 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

–0.8	1	*
–0.6	2	**
–0.4	1	*
–0.2	8	********
0.0	8	********
0.2	30	******************************
0.4	35	***********************************
0.6	14	**************
0.8	1	*

(b) Midpoint Count

–1.2	3	***
–1.0	0	
–0.8	4	****
–0.6	9	*********
–0.4	8	********
–0.2	19	*******************
0.0	16	****************
0.2	17	*****************
0.4	8	********
0.6	6	******
0.8	5	*****
1.0	3	***
1.2	0	
1.4	2	**

Figure 4.9 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

−1.4	1	*
−1.2	1	*
−1.0	2	**
−0.8	2	**
−0.6	7	*******
−0.4	11	***********
−0.2	18	******************
0.0	28	****************************
0.2	24	************************
0.4	4	****
0.6	1	*

(b) Midpoint Count

−1.0	2	*
−0.8	2	**
−0.6	1	*
−0.4	8	********
−0.2	11	***********
0.0	18	******************
0.2	17	*****************
0.4	13	*************
0.6	8	********
0.8	7	*******
1.0	5	*****
1.2	3	***
1.4	1	*
1.6	2	**
1.8	1	*

Figure 4.10 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for clean data

Each * represents 2 observations

(a) Midpoint Count

0.0	2	*
0.2	37	*******************
0.4	55	****************************
0.6	6	***

Each * represents 2 observations

(b) Midpoint Count

−1.5	1	*
−1.0	7	****
−0.5	53	**************************
0.0	38	******************
0.5	1	*

Figure 4.11 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

0.0	8	********
0.2	38	**************************
0.4	46	**
0.6	8	********

(b) Midpoint Count

−1.0	1	*
−0.8	0	
−0.6	3	***
−0.4	7	*******
0.2	27	**************************
0.0	19	******************
0.2	29	*****************************
0.4	8	********
0.6	6	******

Figure 4.12 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

−0.6	1	*
−0.4	10	**********
−0.2	24	**********************
0.0	39	***
0.2	23	**********************
0.4	2	**

(b) Midpoint Count

−0.6	1	*
−0.4	3	***
−0.2	7	*******
0.0	26	**************************
0.2	19	******************
0.4	27	**************************
0.6	7	*******
0.8	8	********
1.0	0	
1.2	1	*

Figure 4.13 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for M–estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for contaminated data (γ=5%)

(a) Midpoint Count

–0.6	1	*
–0.4	0	
–0.2	0	
0.0	5	*****
0.2	17	**********
0.4	40	**********
0.6	32	**********
0.8	5	*****

(b) Midpoint Count

–2.0	1	*
–1.6	2	**
–1.2	12	**********
–0.8	32	**********
–0.4	31	**********
0.0	14	**********
0.4	7	*******
0.8	0	
1.2	0	
1.6	1	*

Figure 4.14 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

–1.0	1	*
–0.8	0	
–0.6	1	*
–0.4	0	
–0.2	3	***
0.0	7	*******
0.2	9	*********
0.4	39	**********
0.6	30	**********
0.8	10	**********

(b) Midpoint Count

−1.6	1	*
−1.2	3	***
−0.8	7	*******
−0.4	37	************************************
0.0	34	**********************************
0.4	11	***********
0.8	5	*****
1.2	0	
1.6	1	*
2.0	1	*

Figure 4.15 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

−1.4	1	*
−1.2	0	
−1.0	1	*
−0.8	0	
−0.6	2	**
−0.4	5	*****
−0.2	8	********
0.0	26	**************************
0.2	36	************************************
0.4	17	*************
0.6	3	***

(b) Midpoint Count

−1.2	1	*
−0.8	4	****
−0.4	23	**********************
0.0	39	**
0.4	22	**********************
0.8	7	*******
1.2	1	*
1.6	0	
2.0	2	**

Figure 4.16 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for contaminated data (γ=5%)

(a) Midpoint Count

0.2	4	****
0.4	25	************************
0.6	29	****************************
0.8	32	********************************
1.0	8	********
1.2	2	**

(b) Midpoint Count

–1.6	3	***
–1.2	19	*******************
–0.8	42	**
–0.4	32	********************************
0.0	4	****

Figure 4.17 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

0.0	1	*
0.2	5	*****
0.4	19	*******************
0.6	23	************************
0.8	27	****************************
1.0	21	**********************
1.2	4	****

(b) Midpoint Count

–1.6	1	*
–1.2	2	**
–0.7	29	*****************************
–0.2	42	**
0.3	22	**********************
0.8	4	****

Figure 4.18 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

(a) Midpoint Count

–0.4	1	*
–0.2	3	***
0.0	10	**********
0.2	20	********************
0.4	28	****************************
0.6	28	****************************
0.8	7	*******
1.0	2	**

(b) Midpoint Count

–1.2	2	**
–0.8	15	***************
–0.4	42	**
0.0	28	****************************
0.4	7	*******
0.8	2	**

Figure 4.19 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for M–estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ when $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$ with non–normal contaminating component (γ=5%)

(a) Midpoint Count

–1.0	1	*
–0.8	0	
–0.6	1	*
–0.4	0	
–0.2	6	******
–0.0	10	**********
0.2	29	*****************************
0.4	36	************************************
0.6	16	****************
0.8	1	*

(b) Midpoint Count

–1.0	1	*
–0.8	4	****
–0.6	9	*********
–0.4	9	*********
–0.2	14	**************
0.0	29	*****************************
0.2	15	***************
0.4	8	********
0.6	3	***
0.8	5	*****
1.0	2	**
1.2	0	
1.4	1	*

Figure 4.20 contaminating component has t_2 distribution

(a) Midpoint Count

–1.4	1	*
–1.2	0	
–1.0	0	
–0.8	2	**
–0.6	2	**
–0.4	2	**
–0.2	5	*****
0.0	12	************
0.2	24	************************
0.4	33	*********************************
0.6	18	******************
0.8	1	*

(b) Midpoint Count

−2.0	1	*
−1.5	0	
−1.0	2	**
−0.5	31	********************************
0.0	33	*********************************
0.5	22	**********************
1.0	6	******
1.5	4	****
2.0	1	*

Figure 4.21 contaminating component has Cauchy distribution

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ when $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$ with non–normal contaminating component (γ=5%)

1 observation below the first class

(a) Midpoint Count

−0.0	5	*****
0.2	33	*********************************
0.4	49	**
0.6	12	************

1 observation above the last class

(b) Midpoint Count

−0.8	1	*
−0.6	2	**
−0.4	11	***********
−0.2	22	**********************
0.0	33	*********************************
0.2	20	********************
0.4	9	*********
0.6	0	
0.8	1	*

Figure 4.22 contaminating component has t_2 distribution

3 observations below the first class

(a) Midpoint Count

−0.2	1	*
0.0	2	**
0.2	20	**********
0.4	45	**********
0.6	11	**********
0.8	4	****
1.0	3	***
1.2	4	****
1.4	1	*
1.6	1	*
1.8	2	**
2.0	2	**
2.2	0	
2.4	1	*

3 observations above the last class

(b) Midpoint Count

−2.5	1	*
−2.0	0	
−1.5	7	*******
−1.0	7	*******
−0.5	24	**********
0.0	46	**********
0.5	11	**********
1.0	1	*

Figure 4.23 contaminating component has Cauchy distribution

Calibration term for $\psi(\sigma_\epsilon, t)$, c_1 , of Median estimators

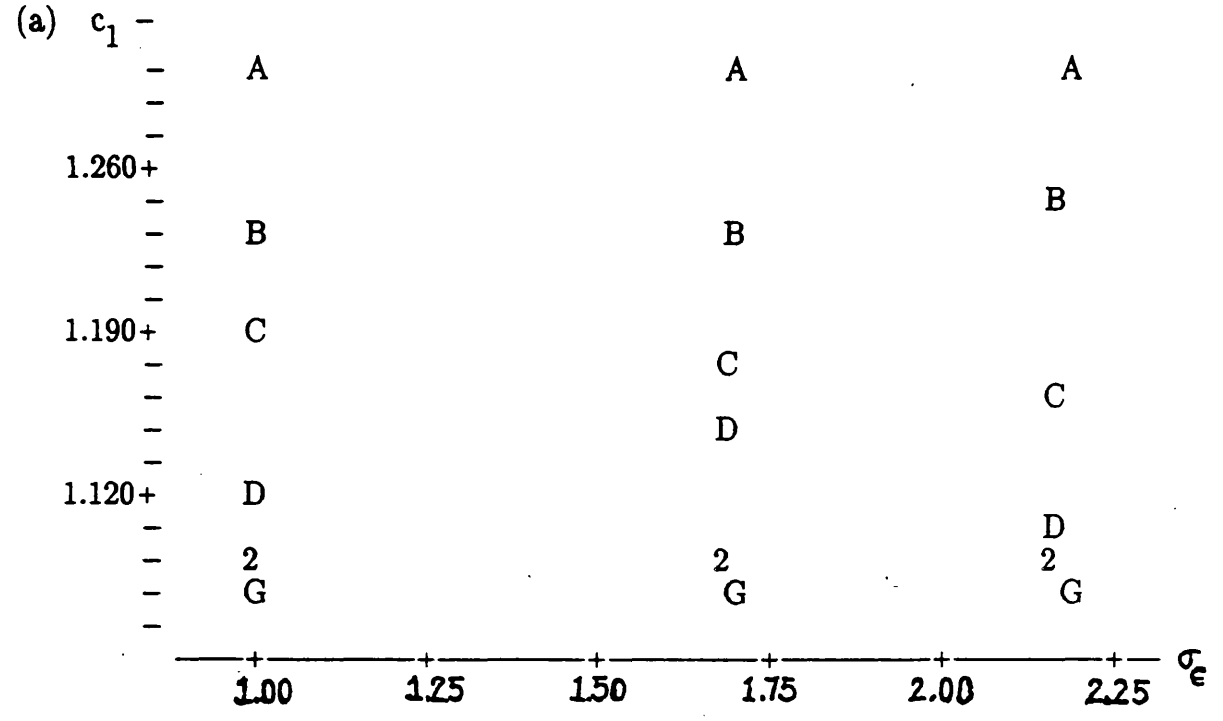


Figure 4.24a Plot of c_1 against σ_ϵ

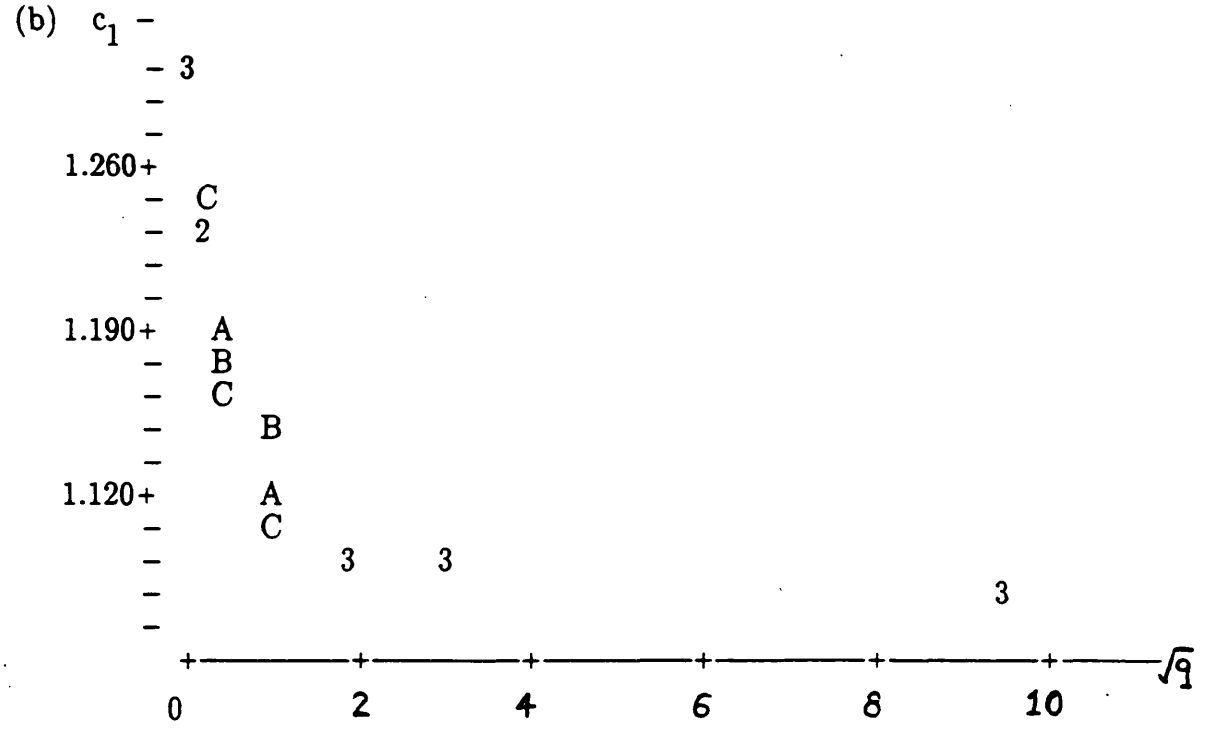


Figure 4.24b Plot of c_1 against $\sqrt{q}$

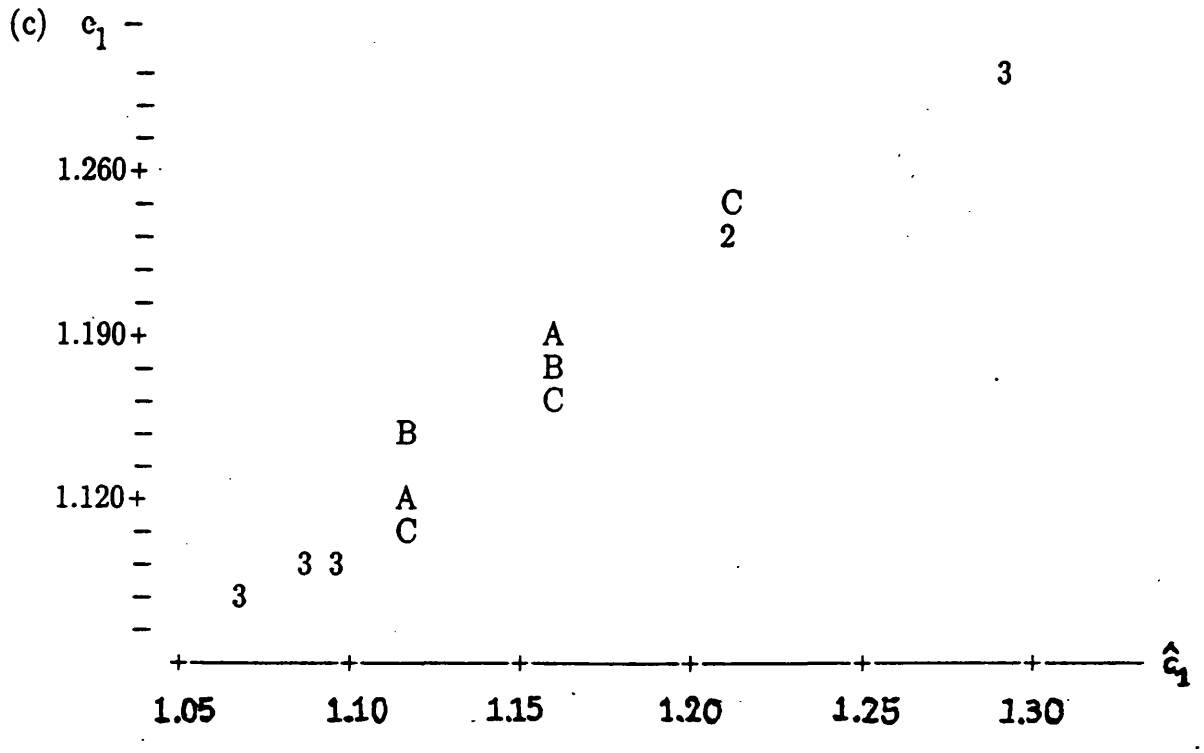


Figure 4.24c Plot c_1 against its fitted values

Calibration term for $\psi(\sqrt{q},t)$, c_2 , of Median estimators

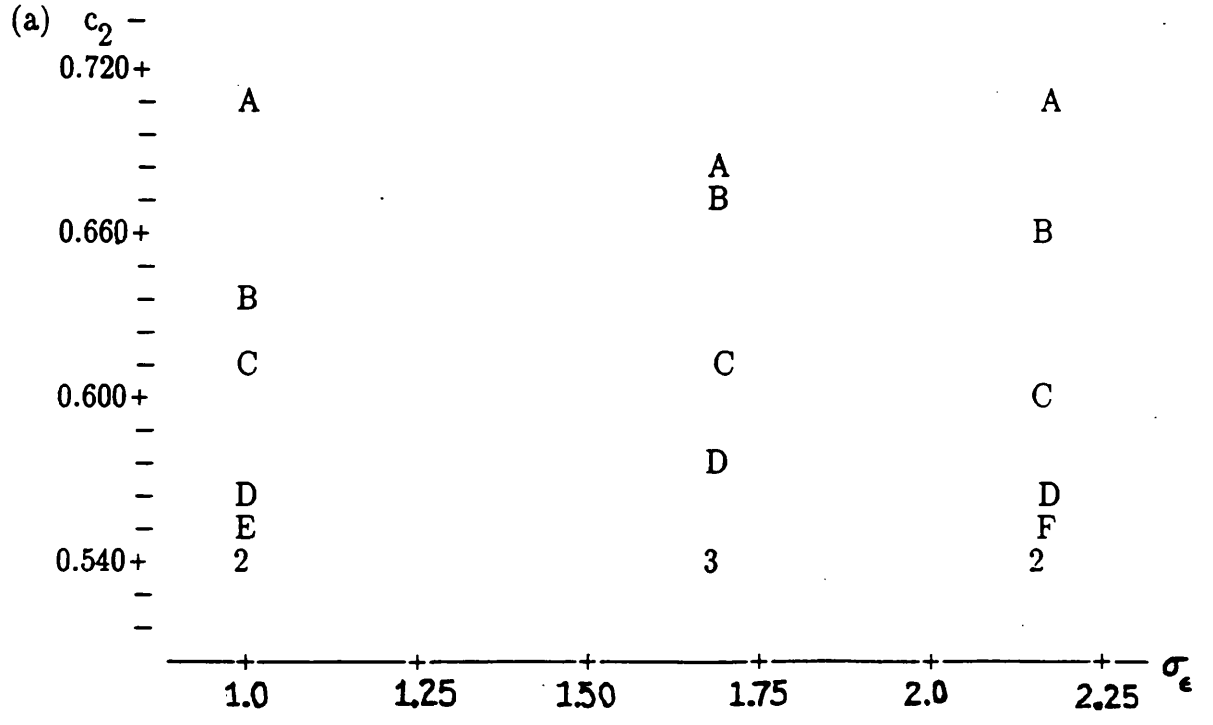


Figure 4.25a Plot c_2 against σ_ϵ

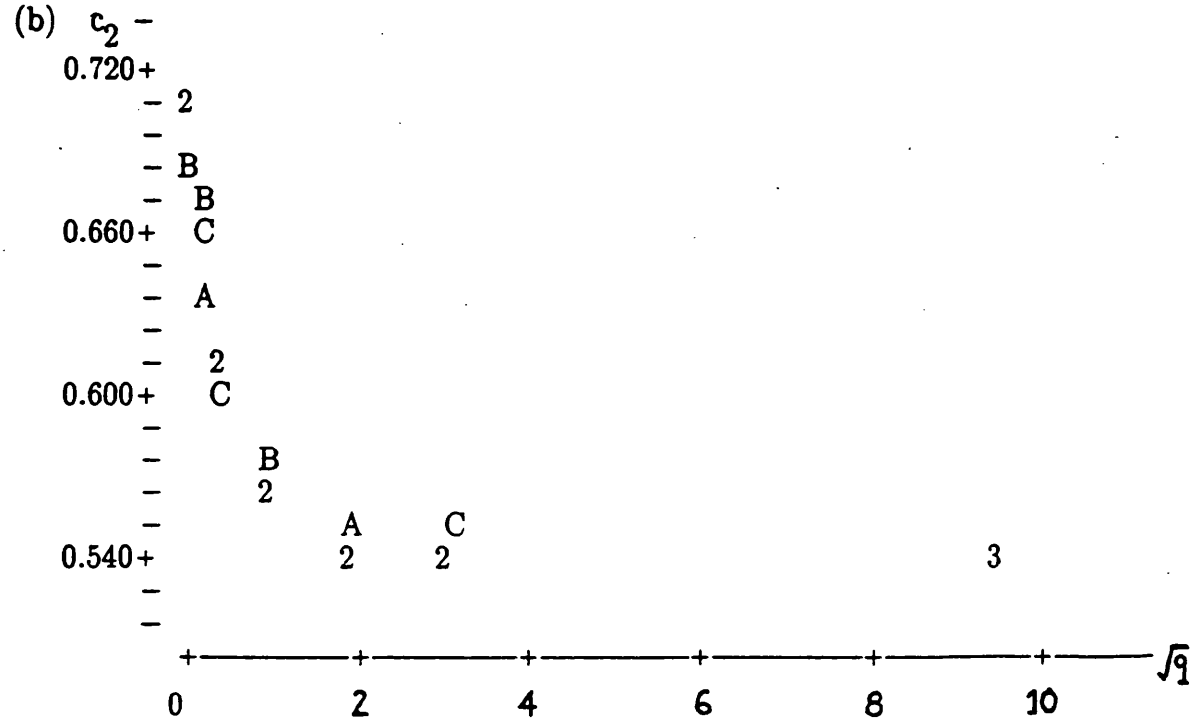


Figure 4.25b Plot of c_2 against $\sqrt{q}$

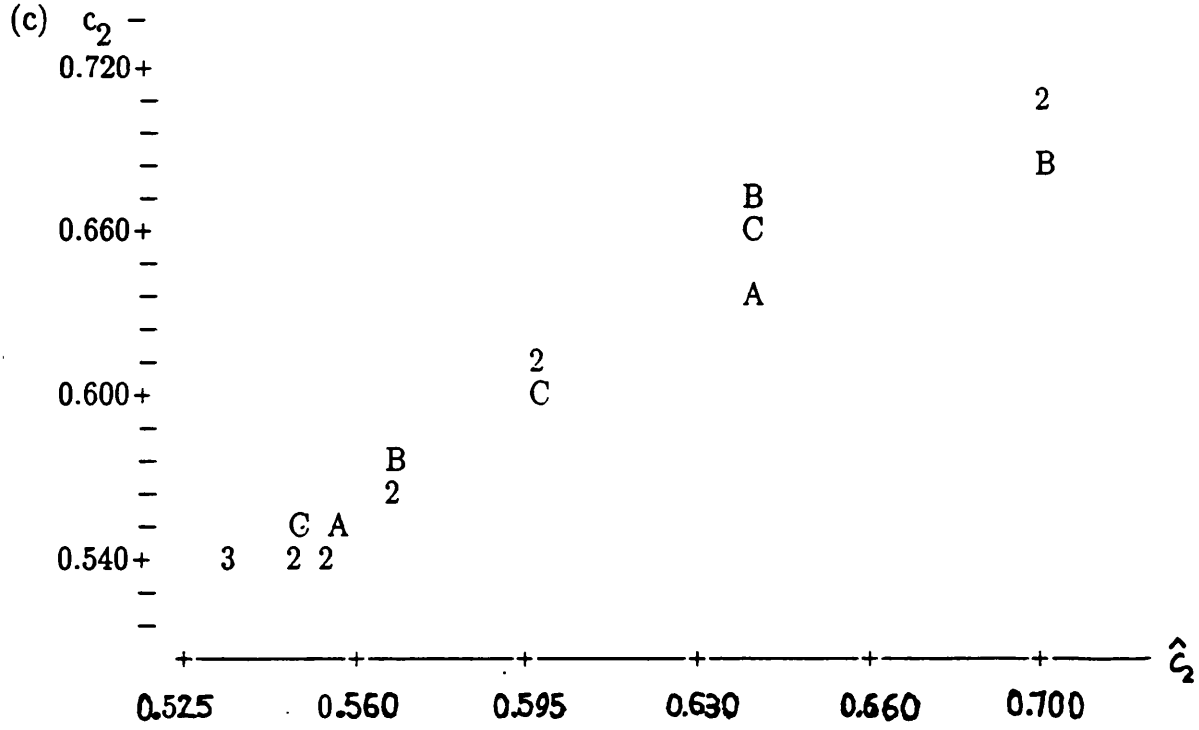


Figure 4.25c Plot of c_2 against its fitted values

Sampling distributions for Median estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for clean data

(a) Midpoint Count

−1.0	1	*
−0.8	0	
−0.6	1	*
−0.4	5	*****
−0.2	3	***
0.0	18	******************
0.2	26	**************************
0.4	33	**********************************
0.6	12	************
0.8	1	*

(b) Midpoint Count

−2.5	1	*
−2.0	1	*
−1.5	4	****
−1.0	12	************
−0.5	31	********************************
0.0	35	************************************
0.5	9	*********
1.0	6	******
1.5	1	*

Figure 4.26 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

1 observation below the first class

(a) Midpoint Count

−0.6	2	**
−0.4	5	*****
−0.2	6	******
0.0	21	**********************
0.2	26	**************************
0.4	25	*************************
0.6	13	*************
0.8	1	*

(b) Midpoint Count

−1.0	4	****
−0.5	24	************************
0.0	29	*****************************
0.5	31	********************************
1.0	10	**********
1.5	1	*
2.0	0	
2.5	1	*

Figure 4.27 $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$

(a) Midpoint Count

−2.2	1	*
−2.0	1	*
−1.8	2	**
−1.6	0	
−1.4	1	*
−1.2	1	*
−1.0	2	**
−0.8	4	****
−0.6	6	******
−0.4	15	***************
−0.2	19	*******************
0.0	21	*********************
0.2	21	*********************
0.4	3	***
0.6	2	**

(b) Midpoint Count

−1.2	1	*
−1.0	2	**
−0.8	2	**
−0.6	1	*
−0.4	6	******
−0.2	12	************
0.0	15	***************
0.2	13	*************
0.4	12	************
0.6	8	********
0.8	8	********
1.0	7	*******
1.2	3	***
1.4	3	***
1.6	1	*
1.8	0	
2.0	1	*
2.2	1	*
2.4	2	**
2.6	1	*

Figure 4.28 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for Median estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for contaminated data (γ=10%)

(a) Midpoint Count

−0.6	1	*
−0.4	0	
−0.2	2	**
0.0	5	*****
0.2	10	**********
0.4	29	*****************************
0.6	41	***
0.8	11	***********
1.0	1	*

(b) Midpoint Count

−2.0	2	**
−1.6	2	**
−1.2	15	***************
−0.8	18	******************
−0.4	39	**************************************
0.0	13	*************
0.4	3	***
0.8	7	*******
1.2	1	*

Figure 4.29 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$; $\gamma = 10\%$ and k = 20

(a) Midpoint Count

−1.4	1	*
−1.2	0	
−1.0	1	*
−0.8	0	
−0.6	0	
−0.4	2	**
−0.2	4	****
0.0	4	****
0.2	9	*********
0.4	25	*************************
0.6	34	**********************************
0.8	18	******************
1.0	2	**

(b) Midpoint Count

−2.5	1	*
−2.0	0	
−1.5	2	**
−1.0	6	******
−0.5	35	**********
0.0	34	**********
0.5	11	***********
1.0	7	*******
1.5	3	***
2.0	0	
2.5	1	*

Figure 4.30 $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$; $\gamma = 10\%$ and k = 25

(a) Midpoint Count

−1.0	2	**
−0.8	0	
−0.6	1	*
−0.4	3	***
−0.2	11	***********
0.0	17	**********
0.2	34	**********
0.4	25	**********
0.6	6	******

(b) Midpoint Count

−1.0	1	*
−0.8	1	*
−0.6	6	******
−0.4	9	*********
−0.2	19	**********
0.0	21	**********
0.2	12	**********
0.4	5	*****
0.6	12	**********
0.8	7	*******
1.0	3	***
1.2	1	*
1.4	1	*
1.6	0	
1.8	1	*

Figure 4.31 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$; $\gamma = 10\%$ and k = 30

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for contaminated data (γ=10%)

(a) Midpoint Count

0.4	7	*******
0.6	20	********************
0.8	31	********************************
1.0	33	**********************************
1.2	8	********
1.4	1	*

(b) Midpoint Count

−2.0	1	*
−1.6	13	*************
−1.2	26	**************************
−0.8	41	***
−0.4	16	****************
0.0	3	***

Figure 4.32 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$; $\gamma = 10\%$ and k = 20

(a) Midpoint Count

0.4	4	****
0.6	16	****************
0.8	28	****************************
1.0	31	********************************
1.2	19	*******************
1.4	2	**

(b) Midpoint Count

−2.0	1	*
−1.5	6	******
−1.0	31	********************************
−0.5	45	***
0.0	17	*****************

Figure 4.33 $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$; $\gamma = 10\%$ and k = 25

(a) Midpoint Count

0.0	1	*
0.2	8	********
0.4	24	************************
0.6	26	**************************
0.8	23	**********************
1.0	15	***************
1.2	2	**

(b) Midpoint Count

−1.6	1	*
−1.4	2	**
−1.2	4	****
−1.0	7	*******
−0.8	17	*****************
−0.6	18	******************
−0.4	21	**********************
−0.2	11	***********
0.0	12	************
0.2	5	*****
0.4	1	*

Figure 4.34 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$; $\gamma = 10\%$ and k = 30

Sampling distributions for Median estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ when $\sigma_\epsilon^2 = \sigma_\eta^2$ = 2 with non–normal contaminating component (γ=10%)

(a) Midpoint Count

−0.6	1	*
−0.4	4	****
−0.2	13	*************
0.0	6	******
0.2	28	****************************
0.4	36	************************************
0.6	11	***********
0.8	1	*

(b) Midpoint Count

−0.8	3	***
−0.6	3	***
−0.4	15	***************
−0.2	14	**************
0.0	21	**********************
0.2	11	***********
0.4	12	************
0.6	6	******
0.8	11	***********
1.0	3	***
1.2	0	
1.4	1	*

Figure 4.35 Contaminating component has t_2 distribution

(a) Midpoint Count

−1.4	1	*
−1.2	1	*
−1.0	1	*
−0.8	1	*
−0.6	3	***
−0.4	3	***
−0.2	6	******
0.0	15	***************
0.2	21	*********************
0.4	34	**********************************
0.6	11	***********
0.8	3	***

(b) Midpoint Count

−1.5	1	*
−1.0	5	*****
−0.5	21	*********************
0.0	38	**************************************
0.5	18	******************
1.0	10	**********
1.5	4	****
2.0	2	**
2.5	1	*

Figure 4.36 Contaminating component has Cauchy distribution

Sampling distributions for MLE of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ when $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$ with non–normal contaminating component (γ=10%)

1 observation below the first class

(a) Midpoint Count

−0.2	1	*
0.0	5	*****
0.2	33	*********************************
0.4	37	*************************************
0.6	19	*******************
0.8	2	**
1.0	1	*
1.2	1	*

1 observation above the last class

(b) Midpoint Count

−1.0	1	*
−0.8	3	***
−0.6	7	*******
−0.4	8	********
−0.2	23	**********
0.0	20	**********
0.2	22	**********
0.4	12	**********
0.6	2	**
0.8	1	*

Figure 4.37 contaminating component has t_2 distribution

4 observations below the first class

(a) Midpoint Count

−0.4	1	*
−0.2	0	
0.0	2	**
0.2	6	******
0.4	30	**********
0.6	16	**********
0.8	13	**********
1.0	3	***
1.2	8	******
1.4	6	****
1.6	3	***
1.8	1	*
2.0	2	**
2.2	1	*
2.4	1	*
2.6	3	***

6 observations below the first class

(b) Midpoint Count

−3.0	1	*
−2.5	1	*
−2.0	3	***
−1.5	6	******
−1.0	17	**********
−0.5	27	**********
0.0	32	**********
0.5	6	******
1.0	1	*

Figure 4.38 contaminating component has Cauchy distribution

Calibration term for $\psi(\sigma_\epsilon, t)$, c_1 of Redescending estimators

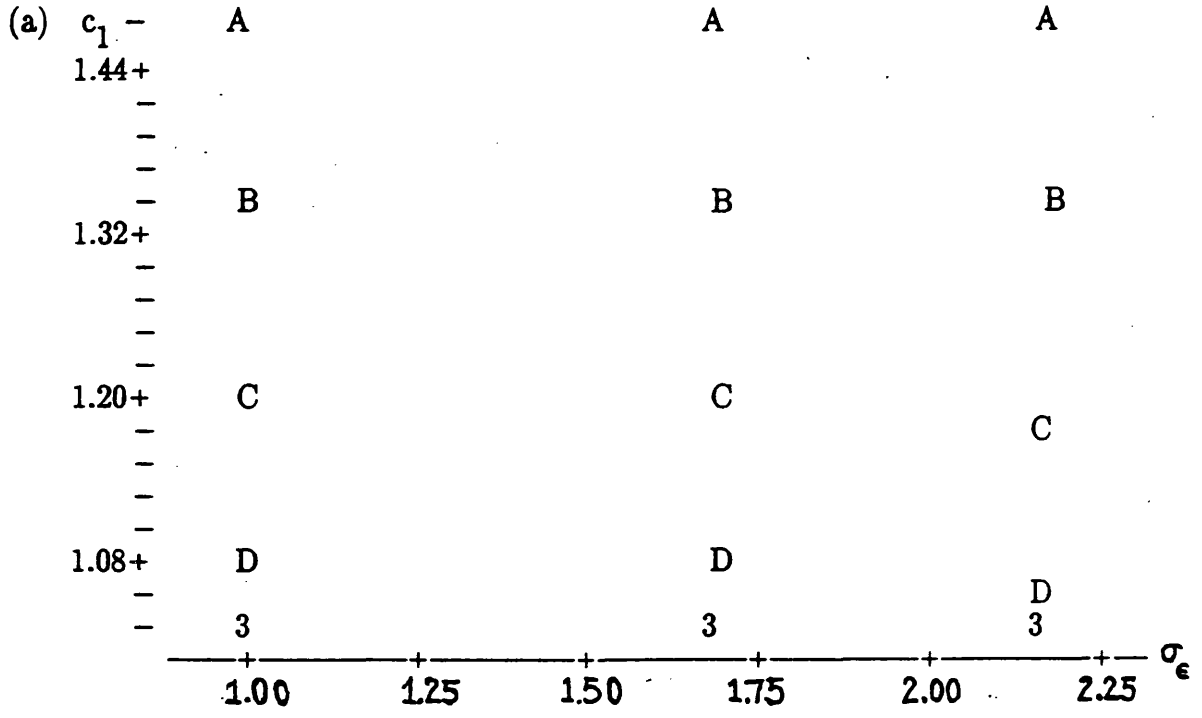


Figure 4.39a Plot of c_1 against σ_ϵ

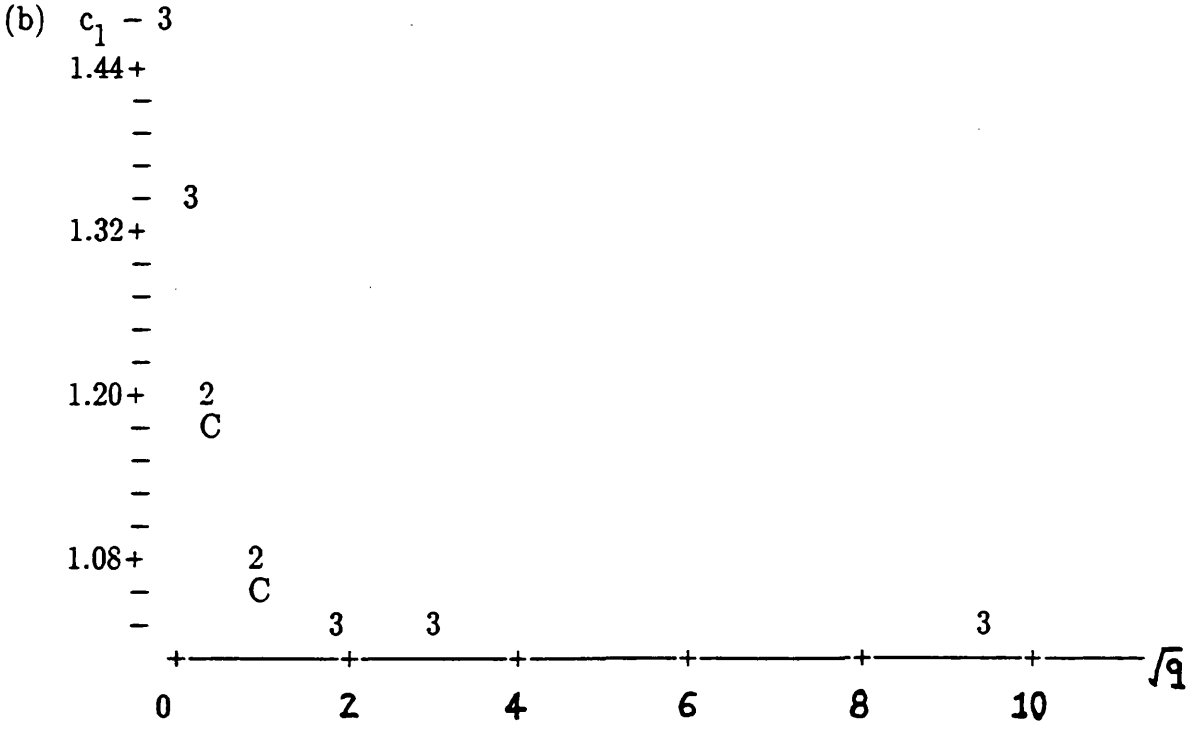


Figure 4.39b Plot of c_1 against $\sqrt{q}$

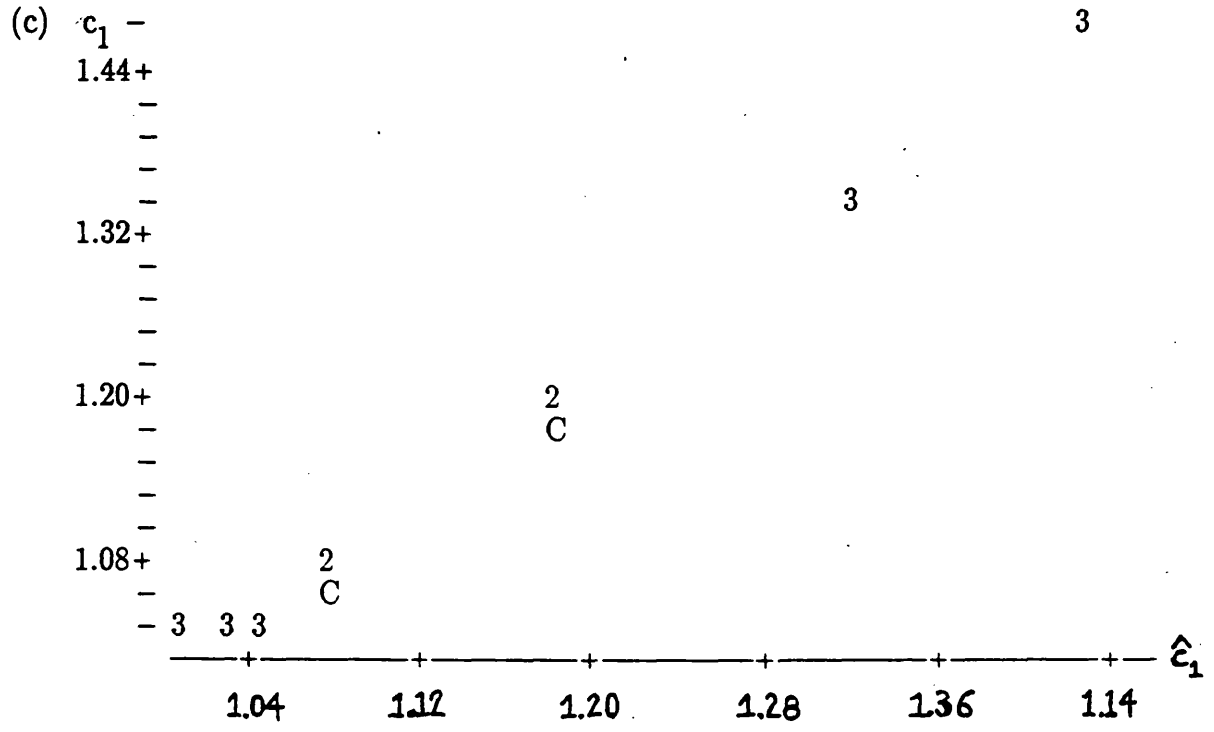


Figure 4.39c Plot of c_1 against its fitted values

Calibration term for $\psi(\sqrt{q},t)$, c_2 of Redescending estimators

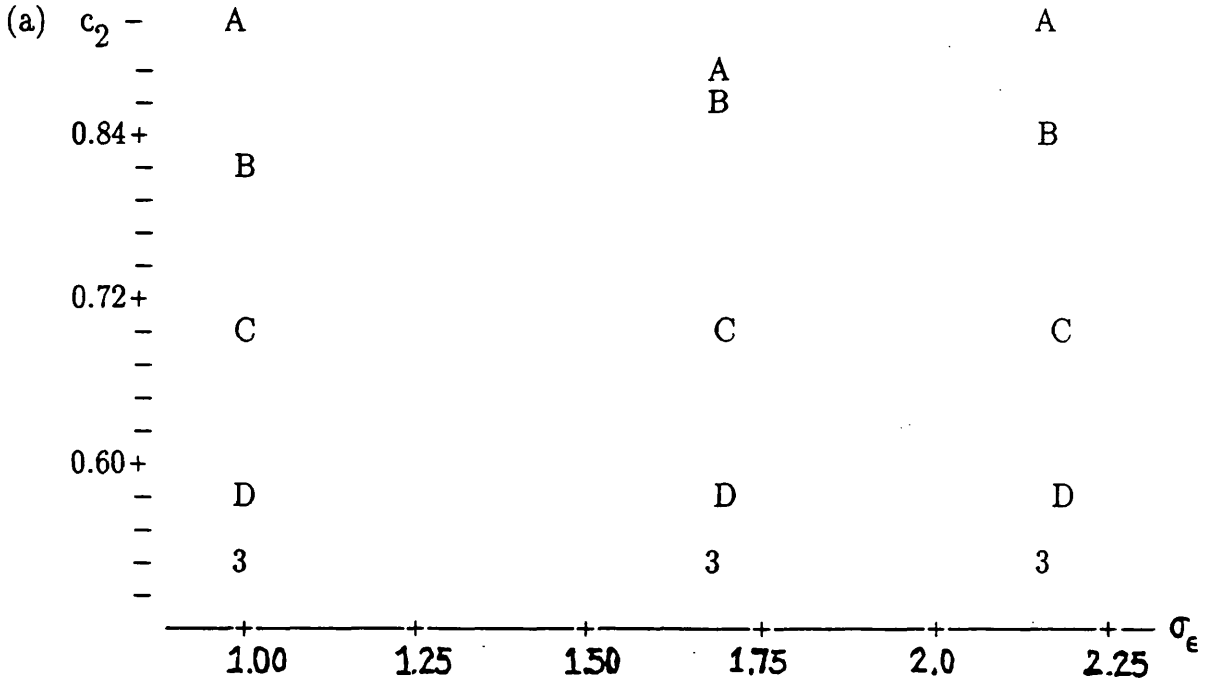


Figure 4.40a Plot of c_2 against σ_ϵ

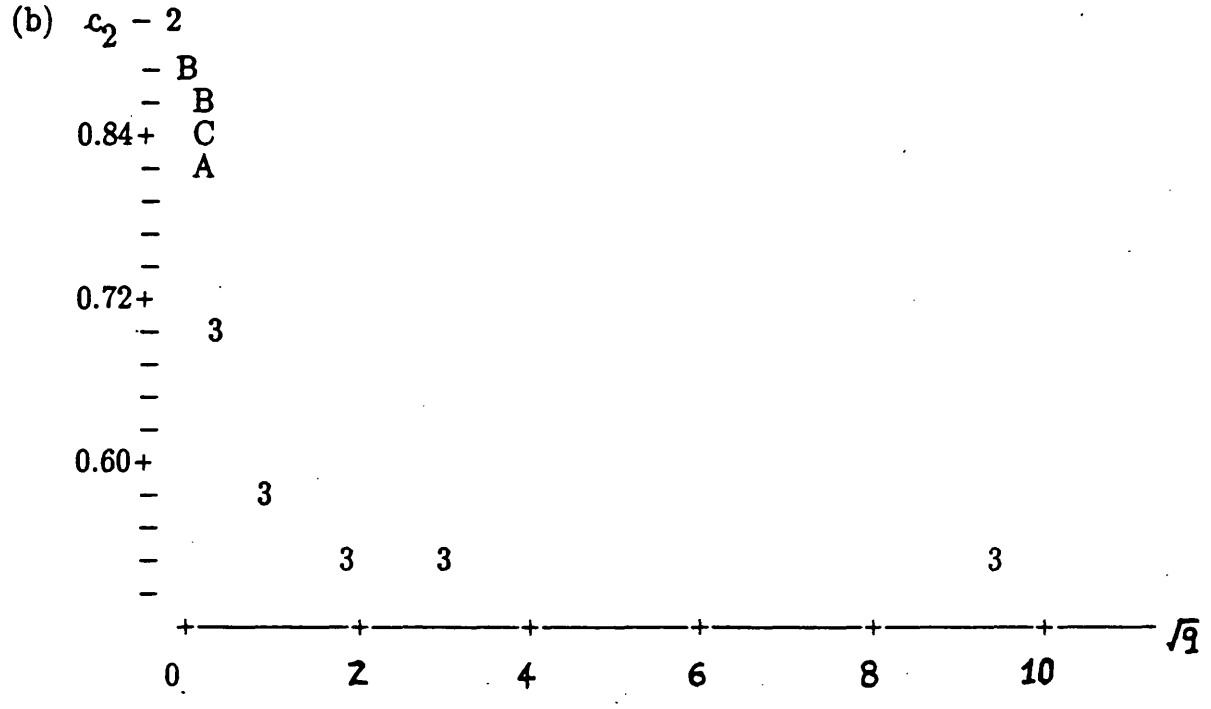


Figure 4.40b Plot of c_2 against $\sqrt{q}$

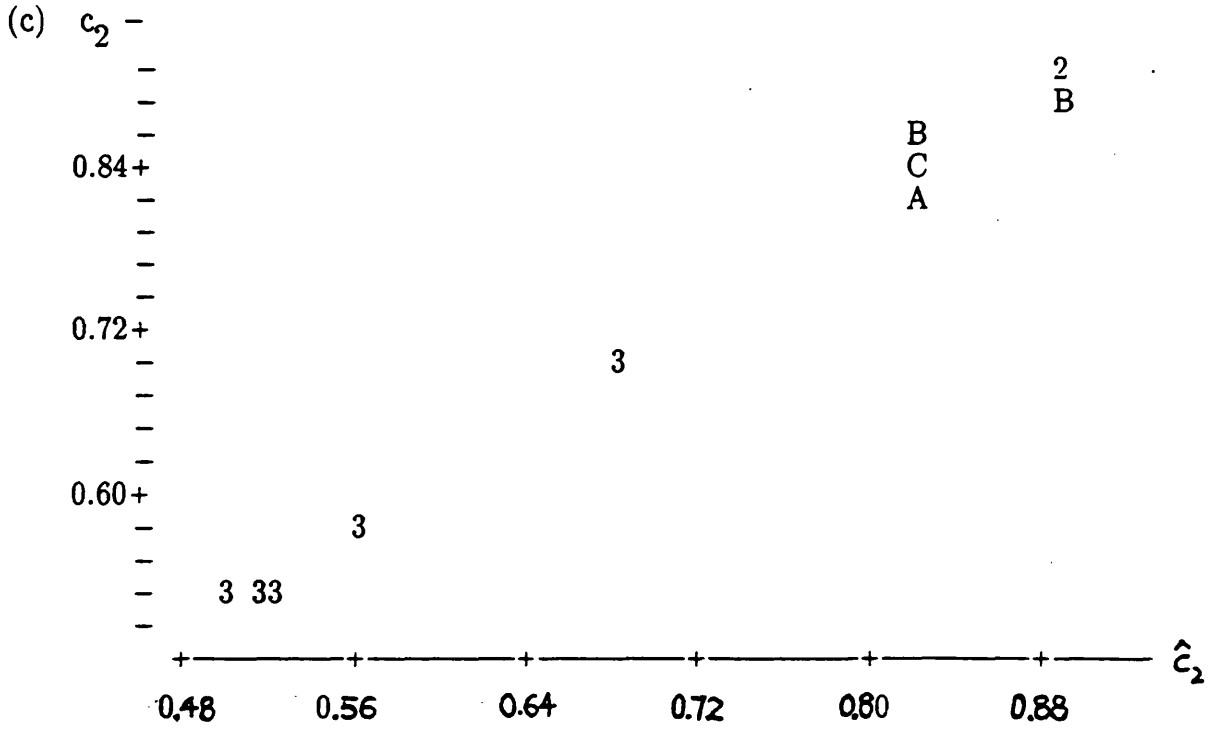


Figure 4.40c Plot of c_2 against its fitted values

Sampling distributions for Redescending estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for clean data

2 observations below the first class

(a) Midpoint Count

−1.4	1	*
−1.2	0	
−1.0	2	**
−0.8	0	
−0.6	4	****
−0.4	3	***
−0.2	8	********
0.0	14	**********
0.2	24	**********
0.4	17	**********
0.6	16	**********
0.8	8	********
1.0	1	*

1 observation above the last class

(b) Midpoint Count

−3.5	3	***
−3.0	4	****
−2.5	3	***
−2.0	7	*******
−1.5	6	******
−1.0	18	**********
−0.5	34	**********
0.0	14	**********
0.5	8	********
1.0	2	**

Figure 4.41 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$

(a) Midpoint Count

−1.4	2	**
−1.2	2	**
−1.0	0	
−0.8	2	**
−0.6	5	*****
−0.4	5	*****
−0.2	4	****
0.0	18	******************
0.2	23	***********************
0.4	15	***************
0.6	17	*****************
0.8	4	****
1.0	3	***

(b) Midpoint Count

−3.5	2	**
−3.0	1	*
−2.5	5	*****
−2.0	6	******
−1.5	8	********
−1.0	8	********
−0.5	18	******************
0.0	30	******************************
0.5	15	***************
1.0	4	****
1.5	2	**
2.0	1	*

Figure 4.42 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$

3 observations below the first class

(a) Midpoint Count

−1.6	3	***
−1.4	0	
−1.2	0	
−1.0	3	***
−0.8	8	********
−0.6	4	****
−0.4	14	**********
−0.2	21	**********
0.0	15	**********
0.2	14	**********
0.4	9	*********
0.6	2	**
0.8	1	*
1.0	2	**

2 observations above the last class

(b) Midpoint Count

−3.5	2	**
−3.0	2	**
−2.5	5	*****
−2.0	3	***
−1.5	6	******
−1.0	9	*********
−0.5	15	**********
0.0	19	**********
0.5	26	**********
1.0	6	******
1.5	3	***
2.0	1	*

Figure 4.43 $\sigma_\epsilon^2 = 1$ and $\sigma_\eta^2 = 1.5$

Sampling distributions for Redescending estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ for contaminated data

2 observations below the last class

(a) Midpoint Count

−1.6	1	*
−1.4	0	
−1.2	1	*
−1.0	2	**
−0.8	1	*
−0.6	1	*
−0.4	2	**
−0.2	5	*****
0.0	12	************
0.2	17	*****************
0.4	26	**************************
0.6	18	******************
0.8	9	*********
1.0	3	***

2 observations above the last class

(b) Midpoint Count

−4.0	1	*
−3.5	4	****
−3.0	3	***
−2.5	3	***
−2.0	10	**********
−1.5	7	*******
−1.0	22	**********************
−0.5	26	**************************
0.0	17	*****************
0.5	5	*****

Figure 4.44 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$; $\gamma = 5\%$ and k = 20

3 observations below the first class

(a) Midpoint Count

−1.6	1	*
−1.4	1	*
−1.2	1	*
−1.0	1	*
−0.8	0	
−0.6	2	**
−0.4	0	
−0.2	7	*******
0.0	7	*******
0.2	18	******************
0.4	29	*****************************
0.6	21	**********************
0.8	7	*******
1.0	1	*
1.2	1	*

1 observation below the first class
1 observation above the last class

(b) Midpoint Count

−3.0	3	***
−2.5	3	***
−2.0	1	*
−1.5	2	**
−1.0	14	**************************
−0.5	27	***************************
0.0	27	***************************
0.5	16	****************
1.0	3	***
1.5	0	
2.0	1	*
2.5	0	
3.0	1	*

Figure 4.45 $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 2$; $\gamma = 5\%$ and k = 25

2 observations below the first class

(a) Midpoint Count

−1.8	1	*
−1.6	1	*
−1.4	1	*
−1.2	1	*
−1.0	0	
−0.8	2	**
−0.6	1	*
−0.4	8	********
−0.2	15	***************
0.0	22	**********************
0.4	28	****************************
0.2	11	***********
0.6	5	*****
0.8	0	
1.0	0	
1.2	1	*

2 observations above the last class

(b) Midpoint Count

−3.6	2	**
−3.2	1	*
−2.8	0	
−2.4	3	***
−2.0	0	
−1.6	3	***
−1.2	4	****
−0.8	9	*********
−0.4	19	*******************
0.0	27	***************************
0.4	21	**********************
0.8	5	*****
1.2	1	*
1.6	1	*
2.0	0	
2.4	1	*

Figure 4.46 $\sigma_{\epsilon}^2 = 1$ and $\sigma_{\eta}^2 = 1.5$; $\gamma = 5\%$ and k = 30

Sampling distributions for Redescending estimator of (a) ln σ_ϵ and (b) ln $\sqrt{q}$ when $\sigma_\epsilon^2 = \sigma_\eta^2 = 2$ with non–normal contaminating component

2 observations below the first class

(a) Midpoint Count

–1.4	1	*
–1.2	2	**
–1.0	0	
–0.8	2	**
–0.6	2	**
–0.4	7	*******
–0.2	6	******
–0.0	9	*********
0.2	18	******************
0.4	27	***************************
0.6	19	*******************
0.8	2	**
1.0	2	**
1.2	1	*

1 observation above the last class

(b) Midpoint Count

–4.0	1	*
–3.5	1	*
–3.0	5	*****
–2.5	1	*
–2.0	2	**
–1.5	3	***
–1.0	6	******
–0.5	17	*****************
0.0	41	***
0.5	13	*************
1.0	7	*******
1.5	2	**

Figure 4.47 contaminating component has t_2 distribution and $\gamma = 5\%$

1 observation below the first class

(a) Midpoint Count

−1.6	1	*
−1.4	0	
−1.2	2	**
−1.0	1	*
−0.8	1	*
−0.6	1	*
−0.4	2	**
−0.2	6	******
0.0	18	******************
0.2	20	********************
0.4	33	*********************************
0.6	9	*********
0.8	1	*
1.0	3	***
1.2	0	
1.4	1	*

1 observation above the last class

(b) Midpoint Count

−3.5	2	**
−3.0	2	**
−2.5	3	***
−2.0	4	****
−1.5	3	***
−2.0	12	************
−0.5	18	******************
0.0	31	*******************************
0.5	16	****************
1.0	5	*****
1.5	3	***

Figure 4.48 contaminating component has Cauchy distribution and $\gamma = 5\%$

Chapter Five

Regressors in Time Series

5.1 Introduction

5.1.1 The Model

We have been developing robust estimation procedures for the local level model . However , in many applications , some variation in the series can be attributed to observable variables . We shall , thus , consider an extension of the local level model to allow for explanatory variables in the system . Keeping the transition equation (4.1.1b) the same , we include regressors in the measurement equation . The model now becomes

$$y_t = \mu_t + \underset{\sim}{x}_t'\underset{\sim}{\delta} + \epsilon_t \qquad (5.1.1a)$$

$$\mu_t = \mu_{t-1} + \eta_t \qquad (5.1.1b)$$

where $\underset{\sim}{x}_t$ is a p×1 vector of explanatory variables
and $\underset{\sim}{\delta}$ is the corresponding p×1 vector of fixed unknown coefficients .

The observable variables are assumed to be weakly exogenous , see Harvey (1989) . This means that they are treated as fixed in repeated samples . Such an assumption allows for the estimation of unknown parameters in the model conditional on the explanatory variables . The regression coefficients $\underset{\sim}{\delta}$ are seldom known in practice and have to be estimated .

5.1.2 State space representation

A state space representation of the above system is achieved by including the regression coefficients in the state vector . Now , we have an augmented state vector $\underset{\sim}{\alpha}_t$ which is $(\mu_t \ \underset{\sim}{\delta}_t')'$ where $\underset{\sim}{\delta}_t = \underset{\sim}{\delta}_0 = \underset{\sim}{\delta}$. Using the time–invariant property of the regression coefficients , that is ,

$$\underset{\sim}{\delta}_t = \underset{\sim}{\delta}_{t-1} , \tag{5.1.2}$$

we can express the model in (5.1.1) as

$$y_t = (1 \ \underset{\sim}{x}_t')\underset{\sim}{\alpha}_t + \epsilon_t \tag{5.1.3a}$$

$$\underset{\sim}{\alpha}_t = \begin{bmatrix} 1 & \underset{\sim}{0}' \\ \underset{\sim}{0} & I \end{bmatrix} \underset{\sim}{\alpha}_{t-1} + \begin{bmatrix} \eta_t \\ \underset{\sim}{0} \end{bmatrix} \tag{5.1.3b}$$

In this way , the regression coefficients $\underset{\sim}{\delta}$ can be estimated along with the state μ_t .

5.2 The Generalised Least Squares Transformation (GLST) Procedure

As an alternative to using the augmented state vector , the estimation of the regression coefficients and other unknown parameters in the model can be carried out by using the GLST procedure , see Ansley and Kohn (1985) . We can view the linear system in (5.1.1) as a regression of y_t on $\underset{\sim}{x}_t$ with correlated errors given by

$$\xi_t = \mu_t + \epsilon_t \tag{5.2.1}$$

A transformation , say L , is performed on this disturbance term to remove its correlation . Such a transformation , which uses the Cholesky decomposition , is equivalent to running the Kalman filter for the stochastic part of the model on the error term and treating the resulting innovations as transformed variable , $L\xi_t$. It follows from the measurement equation that the same transformation , L , has to be carried out on $\underset{\sim}{y} - X\underset{\sim}{\delta}$ where $\underset{\sim}{y}$ is the vector of all the observations , i.e. , $(y_1,y_2,\cdots,y_T)'$ and the T×p matrix X is $(\underset{\sim}{x}_1,\underset{\sim}{x}_2,\cdots,\underset{\sim}{x}_T)'$. Since the Kalman filter is linear , we can run it separately on $\underset{\sim}{y}$ and then on X to produce the prediction errors $L\underset{\sim}{y}$ and LX respectively . The covariance matrix of the transformed variables $L(\underset{\sim}{y} - X\underset{\sim}{\delta})$ is simply that of $L\underset{\sim}{y}$ because the explanatory variables are assumed to be fixed . It is a diagonal matrix F where the diagonal elements are the scale of the innovations f_t produced while filtering $\underset{\sim}{y}$. Our regression model now becomes

$$L\underset{\sim}{y} = LX\,\underset{\sim}{\delta} + L\underset{\sim}{\xi} \tag{5.2.2}$$

where the disturbance term $L\underset{\sim}{\xi}$ is uncorrelated and has covariance matrix F . Conditioned on the other parameters in the model and based on normality assumption , the MLE of the regression coefficient is given by the following weighted least squares estimate

$$\hat{\underset{\sim}{\delta}} = (X'L'F^{-1}LX)^{-1}X'L'F^{-1}L\underset{\sim}{y} \tag{5.2.3}$$

In most cases , the hyperparameters of the model , σ_ϵ^2 and σ_η^2 , are not known and have to be estimated together with the regression coefficients $\underset{\sim}{\delta}$. We note that if the regression coefficients are known , we can treat $y_t - \underset{\sim}{x}_t'\underset{\sim}{\delta}$ as a new series and

apply the Kalman filter for the stochastic part of the model to it . This suggests that we can concentrate the regression parameters out of the likelihood function and use an iterative estimation procedure . Given an initial estimate of the regression coefficients , $\hat{\underset{\sim}{\delta}}_0$, the MLE of the hyperparameters are found by applying the Kalman filter for the local level model on the series $y_t - \underset{\sim}{x}_t'\hat{\underset{\sim}{\delta}}_0$. The Kalman filter based on these hyperparameter estimates is then used to filter $\underset{\sim}{y}$ and X in order to obtain the regression model (5.2.2) . This leads to a new estimate for the regression coefficient $\hat{\underset{\sim}{\delta}}_1$ found by evaluating the expression in (5.2.3) . Using this new estimate of $\underset{\sim}{\delta}$, the hyperparameters are re–estimated . This procedure is repeated until convergence takes place .

The initial estimate $\hat{\underset{\sim}{\delta}}_0$ needed to start the iteration procedure has to be found without involving the hyperparameter values . By taking first differences of the observations in (5.1.1) , we reduce the system to a regression model with a stationary disturbance term . We have

$$\Delta y_t = \Delta \underset{\sim}{x}_t' \underset{\sim}{\delta} + \eta_t + \Delta \epsilon_t \qquad (5.2.4)$$

where the noise term $\eta_t + \Delta\epsilon_t$ follows a MA(1) process . Assuming finite second moments of ϵ_t and η_t , ordinary least squares (OLS) regression of Δy_t on Δx_t produces a consistent though inefficient estimate for $\underset{\sim}{\delta}$ which is used as our $\hat{\underset{\sim}{\delta}}_0$.

5.3 Robustification of the GLST Procedure

The GLST procedure described in the previous section does not allow for the presence of outliers in the series . Neither the scale estimation nor the estimation of the regression coefficients is robust . Suppose we have a series that is from the system defined in (5.1.1) but with a measurement noise term which follows a long–tailed distribution . Application of the GLST procedure on this series will produce parameter estimates which are distorted by additive outliers . To overcome this problem , we can robustify the scale estimation process by replacing the ML estimates from the Kalman filter by M–estimates discussed in Chapter four . When the Kalman filter with robust estimates of the hyperparameters is used to transform the variables $\underset{\sim}{y}$, outliers are carried over to the new variables $L\underset{\sim}{y}$ because the innovations are affected by outlying points . Robust regression technique are , hence, necessary to estimate the coefficients $\underset{\sim}{\delta}$ from these transformed variables . We cannot use a robust filter such as the missing value filter described in Chapter two for the transformation process . This is because most robust filters do not have the following linear property

$$L(\underset{\sim}{y} - X\underset{\sim}{\delta}) = L\underset{\sim}{y} - LX\,\underset{\sim}{\delta} \tag{5.3.1}$$

which is necessary for the iterative estimation procedure .

5.3.1 Regression quantile

Let us begin with the initial estimate of $\underset{\sim}{\delta}$ required to start the iteration

process . In order to accomodate outliers , it is no longer appropriate to use OLS estimates from the reduced model (5.2.4) . This estimate is especially vulnerable to aberrant observations at the beginning or end of the series although those occurring at other parts of the series tend to cancel out . This is because differencing produces outliers of equal magnitude but in opposite directions . Instead , we start the iteration process with a very robust estimate such as the regression median or the .5th regression quantile , see Koenker and Bassett (1978) . A regression quantile is a generalisation of a sample quantile in the location model to the linear model . It can be found by solving the following minimisation problem . Consider a linear model of the form

$$\underset{\sim}{y}^* = X^* \underset{\sim}{\delta} + \underset{\sim}{\xi}^* \tag{5.3.2}$$

where $\underset{\sim}{y}^*$ is $(y_1^*, y_2^*, \ldots, y_T^*)'$, the T×p matrix X^* is $(\underset{\sim}{x}_1^*, \underset{\sim}{x}_2^*, \ldots, \underset{\sim}{x}_T^*)'$ and the noise term ξ_t^* is i.i.d. . Note that we can re–write the system in (5.1.1) in this form by letting $\underset{\sim}{y}^*$ be $L_{\underset{\sim}{y}}.F^{-1/2}$ and X^* be $LX.F^{-1/2}$ where $L_{\underset{\sim}{y}}$, LX and F are the same as those found in (5.2.2) . The θth regression quantile for the model (5.3.2) is defined as the solution to

$$\min_{\underset{\sim}{d}} \sum_{t=1}^{T} \rho_\theta(y_t^* - \underset{\sim}{x}_t^{*\prime}\underset{\sim}{d}) \tag{5.3.3a}$$

where

$$\rho_\theta(u) = \begin{cases} \theta u & u \geq 0 \\ (\theta - 1)u & u < 0 \end{cases} \tag{5.3.3b}$$

for $0 < \theta < 1$

Standard linear programming technique are used to compute the regression quantiles

A program which serves this purpose is given in Koenker and D'Orey (1987) . However , some of the solutions may not be unique . In this case , only a subsequence of $\underset{\sim}{x}_t^*$ is used or the explanatory variable is perturbed to give a unique solution .

5.3.2 Trimmed least squares estimation

Once the iteration procedure is started , we can estimate the regression coefficients at each stage by a more efficient robust method such as trimmed least squares (TLS) estimation , see Welsh (1987) . The latter is also known as the trimmed mean in the linear model . The idea is to identify and remove outlying observations and then compute a least squares estimate based on the remaining data set . In the regression context , an "extreme" observation is usually associated with one that has a large residual . This means that a preliminary estimate is first required to evaluate the residuals of the observations . Then , observations with either large positive or large negative residuals are deleted .

The performance of TLS estimators is dependent on the preliminary estimate used , see Ruppert and Carroll (1980) . A robust preliminary estimate will improve on its efficiency . In the robust GLST procedure , the estimates of the regression coefficients from the previous stage of iteration is used as the preliminary estimate . This is also a TLS estimate except at the first stage where the regression median is used . It follows that the preliminary estimate used at each stage is robust . Given the preliminary estimate , the residual of each transformed observation y_t^* is evaluated . These residuals are then ranked according to their magnitude . Suppose that the trimming proportion is α , observations whose residuals have magnitude at the top $\alpha \times 100\%$, are discarded . We use ten percent trimming for our simulations

below . This level should , however , vary according to how heavily we think the series is contaminated . A higher level of trimming is appropriate when there are more outliers present . But if the proportion is too high , a drop in efficiency will result due to the loss of "good" observations . The regression coefficients are then estimated by the OLS procedure using the rest of the data set .

5.3.3 Influence in regression

In the previous section , we have referred to outlying observations in regression as those with large residuals . There is also a need to consider influential observations , see Atkinson (1986) , when dealing with robust regression . These are points which are crucial to the inferences drawn from the data . Omission of any one influential observation will result in a substantial change in the estimates of the regression coefficients . Since a robust estimate is one which is not overly affected by minor changes in the data set , it is also necessary to discard influential observations. We note that outlying points may not necessarily be influential . This is often the case when the aberrant point is not far enough from the bulk of the observations to cause a "swing" in the fitted regression line . However , it is still necessary to remove such observations in order to reduce the variability of the regression parameters' estimates .

Influential observations are usually leverage points , that is , observations which has an extreme value of one or more explanatory variables . A high leverage point often forces the fitted model close to its observed value leading to a small residual . In order to identify leverage points , let us consider the regression model in (5.3.2) . The "hat" matrix of dimension T×T is defined as

$$H = X^*(X^{*\prime}X^*)^{-1}X^{*\prime} \tag{5.3.4}$$

The diagonal elements of this hat matrix , denoted by h_i , measures the remoteness of an observation in a factor space . Hence , if the observation is a leverage point , the value of h_i will be substantially bigger than the rest of the h_i values . However , not all leverage points are influential observations since they may be in line with the rest of the data . In which case , they may contain important information and their inclusion in the data set will help improve the variability of the estimates . Although we do not trim all leverage points , the measure of leverage is used to identify influential observations .

In order to measure the influence of an observation , we need to evaluate its studentized residual . Ordinary residuals from least squares estimation are given by

$$\underset{\sim}{e} = (I - H)\underset{\sim}{y}^* \tag{5.3.5a}$$

They have a scale dependent distribution with

$$\mathrm{Var}(e_t) = (1 - h_t)\sigma^2 \tag{5.3.5b}$$

for $t = 1,2,...,T$. The value of σ^2 is one in our model (5.3.2) . A studentized version of these residuals has a scale–free distribution and is given by

$$r_t = \frac{e_t}{\sqrt{1 - h_t}} \tag{5.3.6}$$

A statistic D_t for determining whether the tth observation is influential is a function

of both h_t and r_t . We have

$$D_t = \frac{1}{p} r_t^2 \frac{h_t}{1 - h_t} \tag{5.3.7}$$

This is derived based on the difference in the least squares parameter estimates when the tth observation is omitted from the data set , see Cook and Weisberg (1982) . We infer that the ith case is influential if the value of D_i is substantially larger than the bulk of the D_t values .

It is found that if we trim both outlying and influential observations at each stage of the GLST procedure , the iterations become unstable and convergence may be affected . Hence , influential observations are removed only at the last stage and not during the intermediate steps of the iterations . At the final stage of the GLST procedure , residuals computed using the preliminary estimate are ranked according to their magnitude as before . But only those at the top $\frac{2}{3}\alpha \times 100\%$ are discarded . Influential observations based on least squares estimates are identified by evaluating and ranking the statistic D_t for the whole series . Those at the top $\frac{2}{3}\alpha \times 100\%$ are omitted . In this way , the trimming proportion is maintained roughly at the level α. This is because outlying points can sometimes be influential , i.e. , there may be an overlap of points being identified as both aberrant and influential . Thus , the trimming proportion will vary between $\frac{2}{3}\alpha$ and $\frac{4}{3}\alpha$.

With the estimation of regression coefficients and scale being replaced by robust technique , the GLST procedure is made more resistant towards outliers . However , there may be a few cases of non–convergence in the iteration procedure . This can sometimes occur when there is a flip–flop between two sets of parameter

estimates or when the scale estimation is not smoothly carried out , as indicated by a non–zero IFAIL value . It is hence necessary to have an alternative stopping criteria such as fixing the maximum number of iterations . We expect the average number of iterations to increase with the number of exogenous variables . Thus , the maximum number of iterations should be made to depend on the number of explanatory variables in the system , p . As seen in the simulations below such cases of non–convergence usually takes place less than 5% of the time .

To test that this robustified GLST procedure is working properly , we check for uncoupling effects between the explanatory variables X and the regression coefficients $\underset{\sim}{\delta}$. We find that by fixing the observations $\underset{\sim}{y}$ and changing the sign of X , the regression parameter estimates are affected only by a change in sign . Similarly , scaling the explanatory variables by a factor m while holding the observations constant results in the regression coefficient estimates being scaled by $\frac{1}{m}$. Hence , we conclude that there is no uncoupling effect between X and $\underset{\sim}{\delta}$.

A summary of the robustified GLST procedure is given as follows

Step 1 The .5th regression quantile or regression median is computed from the differenced series Δy_t to give us an initial resistant estimate of the regression coefficients $\underset{\sim}{\delta}$.

Step 2 Based on the current estimate of the regression coefficient , robust estimates of the hyperparameters are obtained using the M–estimation technique .

Step 3 Both the series and the explanatory variables are transformed using the new estimates of the hyperparameters . Trimmed least squares estimates are then evaluated from these transformed variables .

Step 4 Iteration between steps 2 and 3 is carried out until convergence takes place or when the maximum number of iterations is reached .

Step 5 The final stage of trimmed least squares estimation of the regression coefficients is repeated . This time both outlying and influential observations are discarded .

5.4 Simulation Results

A comparison of the GLST procedure described in section 5.2 against its robustified version described in the previous section can be carried out by means of Monte Carlo studies . We shall consider series of length a hundred and have two hundred replications for each set of simulations . We use two hundred replications so that the difference found in the summary statistics of the estimators especially those for the regression coefficients are not too dependent on sampling variability . Data is generated from the system (5.1.1) with the outliers modelled by the distribution of the measurement noise term , ϵ_t . As in Chapter four , uncontaminated series are generated with a normal distribution N(0,2) while additive outliers are allowed in the series by adopting a contaminated normal distribution with parameters $\sigma_\epsilon^2 = 2$, $\gamma =$ 5% and k = 100 . A bigger factor of contamination , k = 100 , is used here instead of k = 20 used in previous simulations . This is because outliers are more difficult to detect when explanatory variables are present in the series . The exogenous variables usually increase the range of the series and hence , mask the presence of outliers . Innovation outliers are assumed to be absent from the system and so the transition noise term follows a normal distribution , N(0,2) . Cases with different hyperparameter values are investigated later .

In each set of simulations below , both the GLST procedure and its robustified version are first run on clean data and then on contaminated data . Summary statistics from the simulations consist of the bias and mean square error (MSE) of the estimators . If there is a breakdown in the estimation , i.e. , when the estimate of a parameter in at least one replication is way off from most other estimates within the same simulation , then , the number of cases which broke down , M and a separate set of summary statistics is presented . This more resistant set of statistics , denoted by Bias^* and MSE^* , is derived from the median and pseudo–variance of the sampling distributions , see Chapter four . Sampling distributions of the estimators are also given below .

5.4.1 Non–deterministic exogenous variables

The simulation results are categorized according to the model assumed by the explanatory variables . We begin with only one exogenous variable in the model , p = 1 . Three different cases are considered . They are

Case 1 A stationary process where x_t follows a AR(1) model

$$x_t = 0.5x_{t-1} + \zeta_t \tag{5.4.1}$$

Case 2 A random walk model

$$x_t = x_{t-1} + \zeta_t \tag{5.4.2}$$

Case 3 A random walk plus drift model

$$x_t = x_{t-1} + 0.1 + \zeta_t \qquad (5.4.3)$$

where $\zeta_t \sim N(0,0.25)$ in all the above cases .

In order to study the effect of outliers , it is pertinent to keep the variation of the noise term in the regressor low . The value of $var(\zeta_t)$ is kept small relative to the hyperparameter values to prevent the explanatory variable from dominating the series rendering the effect of outliers insignificant . Otherwise , so much information about the exogenous variable is available that the regression coefficient is well estimated despite the presence of outliers . On the other hand , the hyperparameter values cannot be so big as to completely mask the observable variable . For each simulation , the explanatory variable is generated once and used in all the two hundred replications . The parameter value , δ is fixed at one . Other values will be considered later .

Case 1 Explanatory variable from an AR(1) process

TABLE 5.1 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE
(i) Robust estimator of			
(a)	δ	–0.0734	0.3347
(b)	$\ln \sigma_\epsilon$	–0.1058	0.1032
(c)	$\ln \sigma_\eta$	–0.0623	0.0875
(ii) MLE estimates of			
(a)	δ	–0.0489	0.1812
(b)	$\ln \sigma_\epsilon$	–0.0354	0.0280
(c)	$\ln \sigma_\eta$	–0.0177	0.0405

The sampling distributions of the robust estimators and the MLEs are found in figures 5.1 and 5.2 respectively .

TABLE 5.2 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ	–0.0790	0.3284			
(b)	$\ln \sigma_\epsilon$	0.0250	0.0894			
(c)	$\ln \sigma_\eta$	–0.0167	0.1325			
(ii) MLE of						
(a)	δ	–0.0649	0.5514			
(b)	$\ln \sigma_\epsilon$	0.7349	0.6855			
(c)	$\ln \sigma_\eta$	–0.1184	1.0803	–0.0233	0.0591	1

The sampling distributions of the robust estimators and the MLEs are found in figures 5.3 and 5.4 respectively .

Case 2 Explanatory variable from a random walk model

TABLE 5.3 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE
(i) Robust estimator of			
(a)	δ	–0.0443	0.3309
(b)	$\ln \sigma_\epsilon$	–0.0976	0.0996
(c)	$\ln \sigma_\eta$	–0.0780	0.0926
(ii) MLE estimates of			
(a)	δ	–0.0299	0.1685
(b)	$\ln \sigma_\epsilon$	–0.0264	0.0262
(c)	$\ln \sigma_\eta$	–0.0261	0.0406

The sampling distributions of the robust estimators and the MLEs are found in figures 5.5 and 5.6 respectively .

TABLE 5.4 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ	–0.0533	0.3082			
(b)	$\ln \sigma_\epsilon$	0.0324	0.0933			
(c)	$\ln \sigma_\eta$	–0.0005	0.1249			
(ii) MLE of						
(a)	δ	–0.0615	0.4088			
(b)	$\ln \sigma_\epsilon$	0.7431	0.6966			
(c)	$\ln \sigma_\eta$	–0.4196	5.1133	–0.0426	0.0699	5

The sampling distributions of the robust estimators and the MLEs are found in figures 5.7 and 5.8 respectively .

Case 3 Explanatory variable from a random walk plus drift model

TABLE 5.5 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE
(i) Robust estimator of			
(a)	δ	–0.0280	0.2528
(b)	$\ln \sigma_\epsilon$	–0.0798	0.0571
(c)	$\ln \sigma_\eta$	–0.0647	0.1058
(ii) MLE estimates of			
(a)	δ	–0.0374	0.1697
(b)	$\ln \sigma_\epsilon$	–0.0284	0.0268
(c)	$\ln \sigma_\eta$	–0.0242	0.0414

The sampling distributions of the robust estimators and the MLEs are found in figures 5.9 and 5.10 respectively .

TABLE 5.6 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ	–0.0618	0.3285			
(b)	$\ln \sigma_\epsilon$	0.0387	0.1095			
(c)	$\ln \sigma_\eta$	–0.0287	0.1804			
(ii) MLE of						
(a)	δ	–0.0776	0.3844			
(b)	$\ln \sigma_\epsilon$	0.7404	0.6916			
(c)	$\ln \sigma_\eta$	–0.1378	1.0898	–0.0374	0.0587	1

The sampling distributions of the robust estimators and the MLEs are found in figures 5.11 and 5.12 respectively .

Four factorial replicates are carried out to check the simulation results for different sets of parameter values . They are

TABLE 5.7 FACTORIAL REPLICATES OF SIMULATION FOR ONE EXPLANATORY VARIABLE

	δ	σ_ϵ^2	σ_η^2
Test 1	1	2	2
Test 2	10	1	2
Test 3	1	1	1
Test 4	10	2	1

All four experiments reveal a similar pattern of behaviour for the estimators from the two estimation procedures . Let us first consider the estimation of the regression coefficient across the three cases . In the absence of outliers , the GLST procedure seems to be insensitive to the different models which generate the explanatory variable . Let us denote the mean square error associated with the estimate of the regression coefficient by τ . Hence , the value of τ is very similar in all three cases . In order to make the different cases comparable , we have held everything else constant . Thus , the same parameter values for δ , σ_ϵ^2 , σ_η^2 and $\text{var}(\zeta_t)$ are used to generate data for the three cases . There is , however , no uniformity in the performance of the GLST procedure when outliers are present in the series . The system with a stationary exogenous variable is most affected by contamination . Let the τ value be τ_0 when the data is clean and τ_1 when it is not . Consider the ratio of τ_0 to τ_1 , denoting it by r . This ratio is less than one because of the worsening effect outliers have on the estimates of the regression coefficient . We observe that the r value is smallest in case 1 and increases as we move to cases 2 and 3 . This means that the further we move from stationarity in the explanatory variable , the less sensitive the GLST procedure becomes towards aberrant

observations .

On the other hand , the robust procedure shows similar performance regardless of the model which generates the explanatory variable . Hence , the τ value is approximately the same for all three cases for both contaminated as well as clean data . Besides , the ratio r is nearly one for all the three models . This indicates that the modified GLST procedure is robust towards outliers for all three cases . We , therefore , conclude that this new version of the GLST is resistant to changes .

We now compare the performance of the estimators from the two procedures for the cases of clean and contaminated data . Considering the estimator of the regression coefficient $\hat{\delta}$ only , we generally observe a very small downward bias . No significant difference can be detected between the bias of the robust and the MLEs for both clean and uncontaminated data . This means that both estimators are fairly unbiased regardless of the presence of outliers in the series . Any comparison between the two estimators has to be made by looking at their mean square errors . For clean data , the mean square error of the MLE is smaller than that of the robust estimator . When the series is contaminated , the mean square error of the MLE is bigger than that of the robust estimator in all three cases . This difference turns out to be most significant in case 1 , where the explanatory variable comes from a stationary process . However , there is no significant difference in case 3 where the MLE is not overly affected by outliers . A significant difference for this case can be obtained by increasing the level of contamination . Hence , we conclude that the estimator of the regression coefficient from the robust procedure performs better than that obtained from the GLST procedure in the presence of outliers .

However , we observe that this improvement is not very significant . Firstly ,

the initial estimate for the iteration process $\hat{\delta}_0$ produced by the GLST procedure is not badly affected by the outliers . This is because when we difference the series , an outlier in the original data will appear twice and in opposite directions after transformation . Exceptions to this are cases where discordant values appear at the start or end of the series . Application of OLS estimation on the differenced series will mostly result in a nullifying effect of the outliers on the parameter estimate . Besides , at each stage of the iterative procedure , we are using the transformed variable y_t^* instead of the original observations y_t . Each aberrant value in the original series perturbs several innovations . This results in a smearing effect of the outliers during the transformation process . In addition , outlying points on both sides of the regression line tend to average out during the OLS estimation . Hence , we expect more significant improvement to be made in terms of the scale of the regression coefficient . We shall discuss special configurations of outliers whereby more substantial improvement in the regression parameter estimate is made by the robustified GLST procedure in section 5.4.5 .

With regard to the scale estimators , both the bias and the mean square error of the MLEs are significantly smaller than those of the M–estimators for clean data . We expect a better performance from the MLE when there is no contamination . However , we see a very significant improvement in the both summary statistics when we replace the MLE of $\ln\sigma_\epsilon$ by a M–estimator of scale in the contaminated case . Besides , there is a tendency for the MLE of $\ln\sigma_\eta$ to break down when outliers are present . If we exclude these cases , the MLE of $\ln\sigma_\eta$ is better than the robust estimator even with contamination . This is because the outliers are only present at the measurement equation , thereby more directly affecting the disturbance term ϵ_t . Thus , the MSE^* of the MLE of $\ln\sigma_\eta$ is smaller than the MSE of the robust estimator in the absence of innovation outliers . As mentioned in chapter four , there is a

misspecification of the model when the GLST procedure tries to produce MLE under contamination . We also observe that the scale estimators are not too sensitive towards the model assumed by the exogenous variable .

5.4.2 A deterministic trend

Let us now consider the system (5.5.1) but with a deterministic explanatory variable . In the three cases above , the exogenous variable has been generated with a perturbation term ζ . Suppose now that the explanatory variable takes the form of a deterministic trend , that is , $x_t = t$. This particular model is often used in econometric models . Recall that in both the GLST procedure and its robustified version , transformation of the variables $\underset{\sim}{y}$ and $\underset{\sim}{x}$ is carried out using the Kalman filter . However , when the Kalman filter is run on a deterministic trend t , the prediction errors become constant after a while . This will result in the transformed variable x_t^* having the same value most of the time . A regression plot for this case is found in figure 5.13 below .

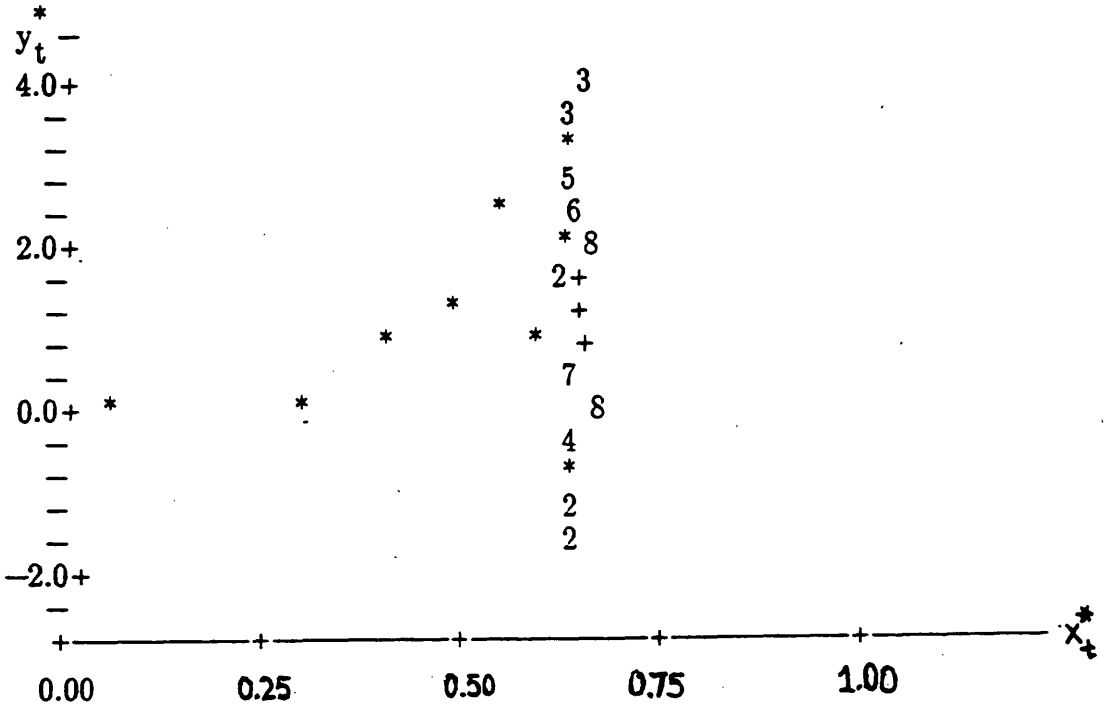


Figure 5.13 Regression plot for the case $x_t = t$

It is clear from the diagram that we do not have very much information on the regression parameter δ . Since the GLST procedure uses all the observations , it is not surprising to see in Tables 5.8 and 5.9 below that the mean square error of the MLE of the regression coefficient is significantly smaller than the robust one for both clean and contaminated data . After transformation , most outliers will fall on either side of the vertical line , rendering the trimming process unnecessary . Again , the bias of the estimators for this regression parameter is rather small . Similar conclusions can be drawn with regards to the scale estimators $\ln\sigma_\epsilon$. That is , the MLE have smaller bias and mean square errors than the M–estimators for clean data whilst the M–estimators have an definite edge over the MLE for contaminated data . As for the other scale estimator $\ln\sigma_\eta$, the MLE starts to break down even when the data is clean . The number of breakdown cases increasing when there is contamination . This pattern of behaviour for the estimators is observed for all the

factorial replicates given in Table 5.7 . As before , we only present the simulation results for Test 1 below .

TABLE 5.8 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ	0.0098	0.0317			
(b)	$\ln \sigma_\epsilon$	–0.0780	0.0917			
(c)	$\ln \sigma_\eta$	–0.0967	0.1014			
(ii) MLE estimates of						
(a)	δ	0.0024	0.0204			
(b)	$\ln \sigma_\epsilon$	–0.0187	0.0275			
(c)	$\ln \sigma_\eta$	–0.1171	1.0543	0.0200	0.0383	1

The sampling distributions of the robust estimators and the MLEs are found in figures 5.14 and 5.15 respectively .

TABLE 5.9 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ	0.0061	0.0332			
(b)	$\ln \sigma_\epsilon$	0.0394	0.1069			
(c)	$\ln \sigma_\eta$	–0.0156	0.1918			
(ii) MLE of						
(a)	δ	–0.0008	0.0205			
(b)	$\ln \sigma_\epsilon$	0.7911	0.7570			
(c)	$\ln \sigma_\eta$	–0.5986	7.1756	–0.0626	0.0824	8

The sampling distributions of the robust estimators and the MLEs are found in figures 5.16 and 5.17 respectively .

5.4.3 Two uncorrelated explanatory variables

We now consider the system (5.1.1) with more than one explanatory variable . Suppose there are two exogenous variables , one from the stationary process in (5.4.1) and another following a local level model given in (5.4.2) . Our aim is to compare the robust estimators with the MLEs for both cases of clean and contaminated data . Since we are estimating more parameters from series of the same length , we expect the estimators here to be more vulnerable to outliers as compared with the case where there is only one regressor in the model . It turns out that a contamination factor of

k = 50 is enough to reveal the differences between the two estimation procedure . To generalise the results across a range of parameter values , the simulations are repeated eight times in the following way .

TABLE 5.10 FACTORIAL REPLICATES OF SIMULATION FOR TWO EXPLANATORY VARIABLES

	δ_1	δ_2	σ_ϵ	σ_η
Test 1	1	1	1	1
Test 2	10	10	1	1
Test 3	10	1	1	2
Test 4	10	1	2	1
Teat 5	1	10	1	2
Test 6	1	10	2	1
Test 7	1	1	2	2
Test 8	10	10	2	2

These replicates produce a similar pattern of behaviour for the estimators . Therefore, we shall only present the simulation results for Test 7 which is most comparable to Test 1 of Table 5.7 .

TABLE 5.11 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE
(i) Robust estimator of			
(a)	δ_1	0.0179	0.2822
(b)	δ_2	–0.0687	0.2836
(c)	$\ln \sigma_\epsilon$	–0.0886	0.0934
(d)	$\ln \sigma_\eta$	–0.1148	0.1001
(ii) MLE of			
(a)	δ_1	–0.0198	0.1642
(b)	δ_2	–0.0343	0.1927
(c)	$\ln \sigma_\epsilon$	–0.0203	0.0260
(d)	$\ln \sigma_\eta$	–0.0683	0.0485

The sampling distributions of the robust estimators and the MLEs are found in figures 5.18 and 5.19 respectively .

TABLE 5.12 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE
(i) Robust estimator of			
(a)	δ_1	0.0005	0.2782
(b)	δ_2	–0.0812	0.2858
(c)	$\ln \sigma_\epsilon$	–0.0455	0.0903
(d)	$\ln \sigma_\eta$	–0.0947	0.1740
(ii) MLE of			
(a)	δ_1	–0.0466	0.3948
(b)	δ_2	–0.0759	0.4114
(c)	$\ln \sigma_\epsilon$	0.5504	0.4019
(d)	$\ln \sigma_\eta$	–0.1025	0.0975

The sampling distributions of the robust estimators and the MLEs are found in figures 5.20 and 5.21 respectively .

In the absence of contamination , we observe that the relative performance of the two estimation procedure is similar whether there is one or more explanatory variables in the system . All the parameters are well estimated by the GLST procedure for clean data . Their mean square errors are significantly smaller than corresponding values from the robust procedure . The bias of both types of estimators for the regression coefficients are negligible . On the other hand , the bias for the scale estimators are significantly different , those of the MLEs being smaller .

When outliers are present in the series , the reverse is true . That is , all the mean square error of the MLEs are significantly bigger than those of the robust estimators with the exception of the hyperparameter $\ln\sigma_\eta$. With the factor of contamination fixed at 50 , the MLE of $\ln\sigma_\eta$ does not break down . Since the contamination does not occur at the transition equation , its mean square error is actually significantly smaller than that of the robust estimator , although there is no significant difference in the bias . The bias in the estimators of the regression coefficients remain quite small . Only the bias of the robust estimator of $\ln\sigma_\epsilon$ is significantly smaller than that of the MLE . We conclude that the robust procedure has an edge over the GLST procedure at the occurrence of aberrant values . This is more apparent when there more exogenous variables in the model .

5.4.4 Two correlated explanatory variables

In real data , the exogenous variables in the system are usually correlated . We shall investigate the effect which correlation amongst regressors has on the estimators from the GLST and robust procedure . Let us respecify the explanatory variables in section 5.4.3 such that for t = 1,2,...,T

$$x_{1t} = 0.5x_{1t-1} + 0.1x_{2t-1} + \zeta_{1t} \qquad (5.4.4a)$$

$$x_{2t} = x_{2t-1} + 0.1x_{1t} + \zeta_{2t} \qquad (5.4.4b)$$

where $x_{10} = x_{20} = 0$ and $\zeta_{it} \sim N(0,0.25)$ for i=1,2

The exogenous variables are still mainly from the stationary and random walk models

but are now correlated . We repeat the simulations in the previous section , including all eight factorial replicates , on this new system . Again , we obtain similar pattern of behaviour for the estimators across different sets of parameter values . The summary statistics for Test 7 of Table 5.10 are found in Tables 5.13 and 5.14 below .

TABLE 5.13 SUMMARY STATISTICS FOR CLEAN DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ_1	0.0118	0.2807			
(b)	δ_2	–0.0646	0.1914			
(c)	$\ln \sigma_\epsilon$	–0.0731	0.0896			
(d)	$\ln \sigma_\eta$	–0.1415	0.1234			
(ii) MLE of						
(a)	δ_1	–0.0279	0.1652			
(b)	δ_2	–0.0338	0.1171			
(c)	$\ln \sigma_\epsilon$	–0.0145	0.0265			
(d)	$\ln \sigma_\eta$	–0.1411	1.0528	0.0494	0.0383	1

The sampling distributions of the robust estimators and the MLEs are found in figures 5.22 and 5.23 respectively .

TABLE 5.14 SUMMARY STATISTICS FOR CONTAMINATED DATA

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ_1	–0.0239	0.2519			
(b)	δ_2	–0.0387	0.1972			
(c)	$\ln \sigma_\epsilon$	0.0582	0.0798			
(d)	$\ln \sigma_\eta$	–0.0845	0.1561			
(ii) MLE of						
(a)	δ_1	–0.0512	0.3983			
(b)	δ_2	–0.0579	0.1985			
(c)	$\ln \sigma_\epsilon$	0.5586	0.4115			
(d)	$\ln \sigma_\eta$	–0.4556	5.1182	–0.0646	0.0570	5

The sampling distributions of the robust estimators and the MLEs are found in figures 5.24 and 5.25 respectively .

Like the uncorrelated case , the bias of all the estimators of the regression coefficients are very small . For clean data , the mean square errors for the MLE of both the regression parameters are significantly smaller than those of the robust estimators . However , we note that the mean square error for the estimator of the second regression parameter δ_2 is smaller than that of the first . This happens in both estimation procedures . The reason for this is that the range of the second explanatory variable is wider than that of the stationary one . The difference being

more marked when there is correlation between the regressors . Contamination in a series will cause the transformed explanatory variables in the GLST procedure to become smaller . As seen earlier , in section 5.4.1 , a series from a stationary process is more affected by outliers than one from a random walk model . Hence , we see a much smaller mean square error for the MLE of the second regression parameter compared to the first . In fact , for this non–stationary exogenous variable , no significant improvement is made by replacing its MLE by a robust estimator at the contamination level of k = 50 . Although the mean square error for the MLE of the first regression coefficient is significantly bigger than that of the robust estimator at that level of contamination .

With correlation present between the two exogenous variables , estimation of the system becomes more difficult . We see this when the MLE of the hyperparameter $\ln\sigma_\eta$ starts to break down for clean data . Disregarding this particular case , the bias and mean square errors for the MLE of both the scale parameters are significantly smaller for clean data . In the presence of outliers , the bias and mean square error for the MLE of $\ln\sigma_\epsilon$ are very much bigger than the summary statistics for the robust estimators and the MLE of $\ln\sigma_\eta$ breaks down more often . Overall , we conclude that the robust procedure has an edge over the GLST procedure when there is contamination in the series .

5.4.5 Special configurations of outliers

As mentioned earlier in section 5.4.1 , the effect of outliers is smeared during the transformation process for removing autocorrelation in the disturbance term ξ_t . Hence , the GLST procedure is able to produce good estimates of the regression

coefficients despite contamination . However , if outliers occur at the start or end of the series , a substantial improvement is made by the modified GLST procedure . This is illustrated in the following two simulations . We use the clean data generated from the system with two correlated explanatory variables , see section 5.4.4 . In the first simulation , we increase the first observation of each replication by a multiple of ten . That is , there is only one outlier occurring at the start of each series . In the second simulation , we contaminate the last observation instead . Again we increase the original data by a factor ten . Running the GLST procedure and its modified version on these two sets of replications gives rise to the following results .

TABLE 5.15 SUMMARY STATISTICS FOR CONTAMINATION AT FIRST OBSERVATION

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ_1	0.0078	0.2613			
(b)	δ_2	–0.0551	0.1797			
(c)	$\ln \sigma_\epsilon$	–0.1601	0.1570			
(d)	$\ln \sigma_\eta$	0.0197	0.0833			
(ii) MLE of						
(a)	δ_1	–0.2016	0.9130			
(b)	δ_2	0.0097	0.1853			
(c)	$\ln \sigma_\epsilon$	–2.6566	45.2992	0.2684	0.9417	45
(d)	$\ln \sigma_\eta$	0.3584	8.4864	0.7044	1.5949	7

The sampling distributions of the robust estimators and the MLEs are found in figures 5.26 and 5.27 respectively .

The above table reveal a substantial improvement made by the robust GLST procedure in the estimation of the first regression coefficient δ_1 . Both the bias and mean square error for the MLE of this parameter is significantly bigger than corresponding values from the robust procedure . Comparing with Table 14 , we find that the MLE is more affected by outliers here . Like before , there is no significant difference in the two estimators for the second regression coefficient δ_2 . Since the outlier is at the start of the series , it will appear as a single outlying point in the differenced series . This will distort the OLS estimate of the regression parameters which are used as the initial estimates of the iteration procedure . The sampling distributions in figures 5.27c and 5.27d show that ML estimation of the hyperparameters breaks down for this type of contamination . On other hand , the M–estimators of scale do not break down and perform well .

TABLE 5.16 SUMMARY STATISTICS FOR CONTAMINATION AT LAST OBSERVATION

	Estimator	Bias	MSE	Bias*	MSE*	M
(i) Robust estimator of						
(a)	δ_1	0.0009	0.2669			
(b)	δ_2	–0.0522	0.1926			
(c)	$\ln \sigma_\epsilon$	–0.0544	0.0866			
(d)	$\ln \sigma_\eta$	–0.1241	0.1129			
(ii) MLE of						
(a)	δ_1	4.5190	28.5553			
(b)	δ_2	–0.8650	7.8185			
(c)	$\ln \sigma_\epsilon$	–5.9366	104.1166	11.112	280.432	102
(d)	$\ln \sigma_\eta$	0.9934	19.4723	2.0734	6.1614	14

The sampling distributions of the robust estimators and the MLEs are found in figures 5.28 and 5.29 respectively .

From the above table , we can easily see the edge which the modified procedure has over the GLST procedure . All the summary statistics for the MLE of the regression coefficients are significantly bigger than those for the robust estimators. Like the previous simulation , contamination at the end of the series will upset the initial estimate used for the iteration process . Besides , the outlier at the last observation will only affect the last innovation from the Kalman filter . Thus , there is no smearing effect in this case .

ML estimation of the hyperparameters is also very unsatisfactory here . For the scale parameter ln σ_ϵ , more than 50% of the cases broke down . The other scale parameter for the transition equation ln σ_η is also badly estimated by the GLST procedure . The sampling distributions for the M–estimators of scale , see figures 5.28c and 5.28d , show that they do not break down . According to their summary statistics , they perform rather well in the estimation of the hyperparameters . Therefore , we conclude that the modified procedure produces more robust estimates of the parameters than the GLST procedure . This is especially so when the outliers appear at the start or end of the series .

5.5 An application to real data

In this section , we consider the application of the robust GLST procedure on a real data set . The latter is that of Durbin and Watson (1951) on the annual consumption of spirits spanning 1870 to 1938 , i.e. , there are 69 observations . The variables are log consumption of spirits per head (y) , log real income per head (x_1) and log relative price of spirits (x_2) . This data set has been analysed by Ansley and Kohn (1989) . They did not use an iterative procedure for their estimation of the parameters in the model . Instead , both the scale and the regression parameters are estimated together using a concentrated likelihood function , see Ansley and Kohn (1985) . Only the first 60 observations were used in their analysis and the estimates for the regression coefficients with their standard deviation are

$\hat{\delta}_1 = 0.69\ (0.13)$

$\hat{\delta}_2 = -0.97\ (0.07)$

Suppose we run the GLST procedure on the first 60 observations of the data set . We obtain the following similar estimates

$$\hat{\delta}_1 = 0.65\ (0.15)$$

$$\hat{\delta}_2 = -0.92\ (0.08)$$

The hyperparameters , σ^2_ϵ and σ^2_η , are estimated as 1.53 and 1.71 respectively . Figures 5.30a to 5.30c show the regression plot for each transformed exogenous variable , namely , y^*_t against x^*_{1t} and y^*_t against x^*_{2t} as well as a plot of one explanatory variable against the other . The diagrams suggest that there are leverage and influential points in the transformed data set .

In Ansley and Kohn (1989) , a measure of influence for the standard regression model , see section 5.3.3 , is generalised for signal plus noise models . Hence , this measure of influence can be used on the model given in equation (5.1.1) . In the paper , observations 46 to 50 were identified as influential and were discarded during the re–estimation of the parameters . The new estimates are

$$\hat{\delta}_1 = 0.6\ (0.13)$$

$$\hat{\delta}_2 = -0.53\ (0.12)$$

By removing the same observations 46 to 50 from the data set and using the GLST procedure , we obtain these estimates

$$\hat{\delta}_1 = 0.49\ (0.16)$$

$$\hat{\delta}_2 = -0.77\ (0.07)$$

Regression plots for the truncated data set are found in figures 5.31a and 5.31b . Comparing with figures 5.30a and 5.30b , the first two diagrams have a smaller range on the vertical axis . This means that the range of the transformed variable y^*_t is smaller after trimming the data set . However , some outlying points still appear in the plots .

Application of the robustified GLST procedure on the original data set gives rise to the following estimates

$\hat{\delta}_1 = 0.52\ (0.14)$

$\hat{\delta}_2 = -0.73\ (0.10)$

The hyperparameters , σ_ϵ^2 and σ_η^2 , are estimated as 1.42 and 1.48 respectively . During the estimation process , 7 observations are discarded because they are either identified as outlying or influential . They are observations 6,40,46,47,49,53 and 58 . Corresponding regression plots are found in figures 5.32a to 5.32c . We notice from these plots that all the transformed variables have smaller range and there are no apparent influential or outlying points .

Explanatory variable from a AR(1) process

(a) Midpoint Count

−0.4	2	**
−0.2	5	*****
0.0	8	********
0.2	14	**********
0.4	19	**********
0.6	25	**********
0.8	22	**********
1.0	26	**********
1.2	21	**********
1.4	24	**********
1.6	14	**********
1.8	13	**********
2.0	4	****
2.2	3	***

Each * represents 2 observations

(b) Midpoint Count

−1.2	1	*
−1.0	2	*
−0.8	0	
−0.6	3	**
−0.4	8	****
−0.2	8	****
0.0	20	**********
0.2	58	**********
0.4	74	**********
0.6	25	**********
0.8	1	*

Each * represents 2 observations

(c) Midpoint Count

−0.8	1	*
−0.6	2	*
−0.4	6	***
−0.2	9	*****
0.0	26	**********
0.2	49	**********
0.4	59	**********
0.6	39	**********
0.8	9	*****

Figure 5.1 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for clean data

(a) Midpoint Count

−0.2	1	*
0.0	4	****
0.2	5	*****
0.4	22	**********
0.6	23	**********
0.8	36	**********
1.0	36	**********
1.2	36	**********
1.4	15	**********
1.6	11	**********
1.8	7	*******
2.0	4	****

Each * represents 5 observations

(b) Midpoint Count

−0.2	6	**
0.0	16	****
0.2	57	**********
0.4	107	**********
0.6	14	***

Each * represents 2 observations

(c) Midpoint Count

−0.6	1	*
−0.4	0	
−0.2	5	***
0.0	13	*******
0.2	61	**********
0.4	78	**********
0.6	39	**********
0.8	3	**

Figure 5.2 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for clean data

(a) Midpoint Count

−0.4	1	*
−0.2	5	*****
0.0	5	*****
0.2	20	********************
0.4	21	*********************
0.6	22	**********************
0.8	23	***********************
1.0	23	***********************
1.2	26	**************************
1.4	23	***********************
1.6	11	***********
1.8	11	***********
2.0	5	*****
2.2	3	***
2.4	1	*

Each * represents 2 observations

(b) Midpoint Count

−0.8	1	*
−0.6	2	*
−0.4	5	***
−0.2	10	*****
0.0	13	*******
0.2	29	***************
0.4	63	*********************************
0.6	59	*******************************
0.8	18	*********

(c) Midpoint Count

−1.2	1	*
−1.0	1	*
−0.8	2	**
−0.6	2	**
−0.4	6	******
−0.2	4	****
0.0	16	****************
0.2	44	**
0.4	49	**
0.6	41	***
0.8	32	*********************************
1.0	2	**

Figure 5.3 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for contaminated data

3 observations above the last class

(a) Midpoint Count

−1.2	1	*
−1.0	1	*
−0.8	1	*
−0.6	1	*
−0.4	6	******
−0.2	6	******
0.0	8	********
0.2	13	*************
0.4	17	*****************
0.6	18	******************
0.8	20	********************
1.0	30	******************************
1.2	19	*******************
1.4	16	****************
1.6	15	***************
1.8	13	*************
2.0	6	******
2.2	4	****
2.4	1	*
2.6	1	*

(b) Midpoint Count

0.0	1	*
0.2	0	
0.4	15	***************
0.6	21	*********************
0.8	29	*****************************
1.0	37	*************************************
1.2	34	**********************************
1.4	30	******************************
1.6	22	**********************
1.8	10	**********
2.0	1	*

Each * represents 2 observations
1 observation below the first class

(c) Midpoint Count

−0.8	2	*
−0.6	0	
−0.4	3	**
−0.2	7	****
0.0	25	************
0.2	57	*****************************
0.4	61	*******************************
0.6	37	******************
0.8	7	****

Figure 5.4 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for contaminated data

Explanatory variable from a random walk model

Each * represents 2 observations

(a) Midpoint Count

−0.4	4	**
0.0	18	*********
0.4	33	*****************
0.8	44	**********************
1.2	57	*****************************
1.6	30	***************
2.0	13	*******
2.4	0	
2.8	1	*

Each * represents 2 observations

(b) Midpoint Count

−1.2	1	*
−1.0	2	*
−0.8	0	
−0.6	2	*
−0.4	8	****
−0.2	9	*****
0.0	18	*********
0.2	55	****************************
0.4	70	***********************************
0.6	34	*****************
0.8	10	*****

Each * represents 2 observations

(c) Midpoint Count

−0.8	1	*
−0.6	3	**
−0.4	7	****
−0.2	9	*****
0.0	34	*****************
0.2	46	**********************
0.4	58	*****************************
0.6	32	****************
0.8	10	*****

Figure 5.5 Sampling distributions for robust estimators of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for clean data

Each * represents 2 observations

(a) Midpoint Count

```
  0.0    7 ****
  0.4   31 ***************
  0.8   73 ************************************
  1.2   58 ****************************
  1.6   27 *************
  2.0    4 **
```

Each * represents 5 observations

(b) Midpoint Count

```
 -0.2    4 *
  0.0   18 ****
  0.2   52 **********
  0.4  108 *********************
  0.6   18 ****
```

Each * represents 2 observations

(c) Midpoint Count

```
 -0.6    1 *
 -0.4    0
 -0.2    4 **
  0.0   19 **********
  0.2   58 *****************************
  0.4   81 ****************************************
  0.6   35 *****************
  0.8    2 *
```

Figure 5.6 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for clean data

(a) Midpoint Count

–0.2	5	*****
0.0	11	**********
0.2	11	**********
0.4	19	**********
0.6	20	**********
0.8	31	**********
1.0	28	**********
1.2	18	**********
1.4	22	**********
1.6	12	**********
1.8	17	**********
2.0	3	***
2.2	2	**
2.4	1	*

Each * represents 2 observations

(b) Midpoint Count

–0.8	1	*
–0.6	6	***
–0.4	3	**
–0.2	3	**
0.0	16	********
0.2	32	**********
0.4	62	**********
0.6	62	**********
0.8	15	********

(c) Midpoint Count

–1.0	1	*
–0.8	1	*
–0.6	3	***
–0.4	5	*****
–0.2	9	*********
–0.0	20	**********
0.2	42	**********
0.4	45	**********
0.6	43	**********
0.8	29	**********
1.0	2	**

Figure 5.7 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for contaminated data

(a) Midpoint Count

−0.6	2	**
−0.4	6	******
−0.2	3	***
0.0	11	**********
0.2	11	**********
0.4	13	************
0.6	24	************************
0.8	27	***************************
1.0	19	******************
1.2	20	********************
1.4	26	**************************
1.6	17	*****************
1.8	10	**********
2.0	6	******
2.2	3	***
2.4	0	
2.6	1	*
2.8	0	
3.0	1	*

(b) Midpoint Count

0.2	1	*
0.4	12	************
0.6	22	**********************
0.8	31	*******************************
1.0	36	************************************
1.2	35	***********************************
1.4	30	******************************
1.6	22	**********************
1.8	10	**********
2.0	1	*

Each * represents 2 observations
5 observations below the first class

(a) Midpoint Count

−0.8	1	*
−0.6	2	*
−0.4	3	**
−0.2	11	******
0.0	24	************
0.2	51	*************************
0.4	69	**********************************
0.6	28	**************
0.8	6	***

Figure 5.8 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for contaminated data

Explanatory variable from a random walk plus drift model

(a) Midpoint Count

–0.4	1	*
–0.2	8	********
0.0	8	********
0.2	10	**********
0.4	17	*****************
0.6	19	*******************
0.8	30	******************************
1.0	25	*************************
1.2	25	*************************
1.4	25	*************************
1.6	17	*****************
1.8	5	*****
2.0	6	******
2.2	3	***
2.4	1	*

Each * represents 2 observations

(b) Midpoint Count

–1.2	1	*
–1.0	1	*
–0.8	1	*
–0.6	2	*
–0.4	6	***
–0.2	11	******
0.0	19	**********
0.2	56	*****************************
0.4	73	*************************************
0.6	28	**************
0.8	2	*

Each * represents 2 observations

(c) Midpoint Count

–0.8	1	*
–0.6	3	**
–0.4	6	***
–0.2	9	*****
0.0	32	****************
0.2	49	*************************
0.4	58	*****************************
0.6	32	****************
0.8	10	*****

Figure 5.9 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for clean data

(a) Midpoint Count

0.0	3	***
0.2	6	******
0.4	23	**********
0.6	23	**********
0.8	27	**********
1.0	43	**********
1.2	34	**********
1.4	18	**********
1.6	16	**********
1.8	6	******
2.0	1	*

Each * represents 5 observations

(b) Midpoint Count

–0.2	4	*
0.0	18	****
0.2	53	**********
0.4	108	**********
0.6	17	****

Each * represents 2 observations

(c) Midpoint Count

–0.6	1	*
–0.4	0	
–0.2	5	***
0.0	16	********
0.2	60	**********
0.4	82	**********
0.6	34	**********
0.8	2	*

Figure 5.10 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for clean data

(a) Midpoint Count

–0.6	2	**
–0.4	1	*
–0.2	3	***
0.0	11	**********
0.2	13	**********
0.4	14	**********
0.6	16	**********
0.8	30	**********
1.0	35	**********
1.2	23	**********
1.4	23	**********
1.6	7	*******
1.8	12	**********
2.0	7	*******
2.2	2	**
2.4	1	*

Each * represents 2 observations

(b) Midpoint Count

–1.0	2	*
–0.8	0	
–0.6	6	***
–0.4	2	*
–0.2	2	*
0.0	14	*******
0.2	30	**********
0.4	66	**********
0.6	58	**********
0.8	17	*********
1.0	1	*
1.2	2	*

Each * represents 2 observations

(c) Midpoint Count

–2.0	1	*
–1.6	0	
–1.2	3	**
–0.8	3	**
–0.4	10	*****
0.0	45	**********
0.4	84	**********
0.8	53	**********
1.2	1	*

Figure 5.11 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for contaminated data

(a) Midpoint Count

–0.8	1	*
–0.6	1	*
–0.4	3	***
–0.2	4	****
0.0	14	**********
0.2	10	**********
0.4	17	**********
0.6	18	**********
0.8	29	**********
1.0	20	**********
1.2	21	**********
1.4	30	**********
1.6	8	********
1.8	18	**********
2.0	4	****
2.2	1	*
2.4	1	*

(b) Midpoint Count

0.2	1	*
0.4	12	**********
0.6	22	**********
0.8	29	**********
1.0	38	**********
1.2	35	**********
1.4	30	**********
1.6	22	**********
1.8	10	**********
2.0	1	*

Each * represents 5 observations
1 observation below the first class

(c) Midpoint Count

–1.2	1	*
–0.8	1	*
–0.4	11	***
0.0	52	**********
0.4	119	**********
0.8	15	***

Figure 5.12 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for contaminated data

Explanatory variable being a deterministic trend

(a) Midpoint Count

0.5	2	**
0.6	3	***
0.7	11	**********
0.8	23	**********
0.9	31	**********
1.0	50	**********
1.1	36	**********
1.2	26	**********
1.3	13	**********
1.4	4	****
1.5	1	*

Each * represents 2 observations

(b) Midpoint Count

–1.0	1	*
–0.8	2	*
–0.6	2	*
–0.4	6	***
–0.2	7	****
0.0	22	**********
0.2	51	**********
0.4	67	**********
0.6	40	**********
0.8	2	*

Each * represents 2 observations

(c) Midpoint Count

–1.0	1	*
–0.8	1	*
–0.6	1	*
–0.4	8	****
–0.2	11	******
0.0	31	**********
0.2	58	**********
0.4	47	**********
0.6	34	**********
0.8	8	****

Figure 5.14 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for clean data

Each * represents 2 observations

(a) Midpoint Count

0.6	2	*
0.7	6	***
0.8	20	**********
0.9	40	********************
1.0	63	********************************
1.1	36	******************
1.2	24	************
1.3	9	*****

Each * represents 2 observations

(b) Midpoint Count

–0.4	1	*
–0.2	2	*
0.0	10	*****
0.2	63	********************************
0.4	100	**
0.6	24	************

Each * represents 2 observations
1 observation below the first class

(c) Midpoint Count

–0.6	2	*
–0.4	0	
–0.2	6	***
0.0	17	*********
0.2	62	********************************
0.4	85	***
0.6	25	*************
0.8	2	*

Figure 5.15 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for clean data

(a) Midpoint Count

0.3	1	*
0.4	0	
0.5	2	**
0.6	2	**
0.7	7	*******
0.8	23	**********
0.9	38	**********
1.0	48	**********
1.1	39	**********
1.2	23	**********
1.3	9	*********
1.4	8	********

Each * represents 2 observations

(b) Midpoint Count

−0.8	2	*
−0.6	1	*
−0.4	7	****
−0.2	6	***
0.0	14	*******
0.2	33	**********
0.4	62	**********
0.6	49	**********
0.8	22	**********
1.0	2	*
1.2	2	*

Each * represents 2 observations

(c) Midpoint Count

−1.4	3	***
−1.2	0	
−1.0	2	**
−0.8	1	*
−0.6	4	****
−0.4	7	*******
−0.2	9	*********
0.0	22	**********
0.2	32	**********
0.4	40	**********
0.6	45	**********
0.8	27	**********
1.0	8	********

Figure 5.16 Sampling distributions for robust estimator of (a) δ, (b) $\ln \sigma_\epsilon$ and (c) $\ln \sigma_\eta$ for contaminated data

Each * represents 2 observations

(a) Midpoint Count

```
 0.6    1 *
 0.7    7 ****
 0.8   22 **********
 0.9   45 **********************
 1.0   54 ***************************
 1.1   37 ******************
 1.2   26 *************
 1.3    8 ****
```

(b) Midpoint Count

```
 0.2    4 ****
 0.4    7 ******
 0.6   12 ************
 0.8   27 ***************************
 1.0   37 *************************************
 1.2   43 *******************************************
 1.4   36 ************************************
 1.6   25 *************************
 1.8    9 *********
```

Each * represents 2 observations
8 observations below the first class

(c) Midpoint Count

```
-1.2    1 *
-1.0    0
-0.8    1 *
-0.6    3 **
-0.4    2 *
-0.2   12 ******
 0.0   28 **************
 0.2   52 **************************
 0.4   59 *******************************
 0.6   28 **************
 0.8    5 ***
 1.0    1 *
```

Figure 5.17 Sampling distributions for MLE of (a) δ ,(b) ln σ_ϵ and (c) ln σ_η for contaminated data

Two uncorrelated explanatory variables :
x_{1t} from AR(1) process and x_{2t} from a random walk model

(a) Midpoint Count

−0.4	3	***
−0.2	0	
0.0	6	******
0.2	10	**********
0.4	10	**********
0.6	22	**********
0.8	33	**********
1.0	30	**********
1.2	25	**********
1.4	25	**********
1.6	16	**********
1.8	10	**********
2.0	5	*****
2.2	4	****
2.4	1	*

(b) Midpoint Count

−0.6	1	*
−0.4	0	
−0.2	3	***
0.0	6	******
0.2	12	**********
0.4	19	**********
0.6	31	**********
0.8	25	**********
1.0	22	**********
1.2	34	**********
1.4	18	**********
1.6	15	**********
1.8	8	********
2.0	3	***
2.2	2	**
2.4	1	*

(a) Midpoint Count

−1.0	2	**
−0.9	0	
−0.8	1	*
−0.7	2	**
−0.6	1	*
−0.5	0	
−0.4	1	*
−0.3	4	****
−0.2	6	******
−0.1	8	********
0.0	8	********
0.1	16	****************
0.2	24	************************
0.3	44	**
0.4	33	*********************************
0.5	32	********************************
0.6	14	**************
0.7	3	***
0.8	1	*

Each * represents 2 observations

(d) Midpoint Count

−0.8	2	*
−0.6	2	*
−0.4	8	****
−0.2	10	*****
0.0	37	*****************************
0.2	49	*************************
0.4	54	****************************
0.6	30	***************
0.8	8	****

Figure 5.18 Sampling distributions for robust estimator of (a) δ_1 , (b) δ_2 , (c) ln σ_ϵ and (d) ln σ_η for clean data

(a) Midpoint Count

2.0	1	*
0.0	1	*
0.2	5	*****
0.4	17	*****************
0.6	23	**********************
0.8	42	**
1.0	36	************************************
1.2	35	***********************************
1.4	16	****************
1.6	15	***************
1.8	6	******
2.0	3	***

(b) Midpoint Count

0.0	4	****
0.2	8	********
0.4	23	**********************
0.6	20	********************
0.8	31	*******************************
1.0	37	************************************
1.2	30	******************************
1.4	22	**********************
1.6	19	*******************
1.8	4	****
2.0	0	
2.2	1	*
2.4	1	*

Each * represents 2 observations

(c) Midpoint Count

–0.2	1	*
–0.1	6	***
0.0	5	***
0.1	16	********
0.2	26	*************
0.3	47	************************
0.4	56	****************************
0.5	33	*****************
0.6	9	*****
0.7	1	*

Each * represents 2 observations

(d) Midpoint Count

−0.6	1	*
−0.4	0	
−0.2	12	******
0.0	18	*********
0.2	68	**********
0.4	77	**********
0.6	21	**********
0.8	3	**

Figure 5.19 Sampling distributions for MLE estimator of (a) δ_1 , (b) δ_2 , (c) ln σ_ϵ and (d) ln σ_η for clean data

Each * represents 2 observations

(a) Midpoint Count

−0.5	4	**
0.0	10	*****
0.5	43	**********
1.0	82	**********
1.5	46	**********
2.0	14	*******
2.5	1	*

(b) Midpoint Count

−0.4	2	**
−0.2	4	****
0.0	6	******
0.2	16	**********
0.4	15	**********
0.6	27	**********
0.8	29	**********
1.0	29	**********
1.2	26	**********
1.4	15	**********
1.6	14	**********
1.8	12	**********
2.0	4	****
2.2	1	*

Each * represents 2 observations

(c) Midpoint Count

−1.0	1	*
−0.8	2	*
−0.6	0	
−0.4	1	*
−0.2	9	*****
0.0	14	*******
0.2	34	**********
0.4	56	**********
0.6	68	**********
0.8	12	******
1.0	2	*
1.2	1	*

Each * represents 2 observations

(d) Midpoint Count

−1.6	1	*
−1.2	0	
−0.8	8	****
−0.4	16	********
0.0	53	**********
0.4	85	**********
0.8	36	**********
1.2	1	*

Figure 5.20 Sampling distributions for robust estimator of (a) δ_1, (b) δ_2, (c) ln σ_ϵ and (d) ln σ_η for contaminated data

Each * represents two observations

(a) Midpoint Count

–1.5	1	*
–1.0	1	*
–0.5	6	***
0.0	14	*******
0.5	49	************************
1.0	72	************************************
1.5	38	*******************
2.0	12	******
2.5	6	***
3.0	1	****

(b) Midpoint Count

–0.6	2	**
–0.4	1	*
–0.2	7	*******
0.0	10	**********
0.2	15	***************
0.4	24	************************
0.6	14	**************
0.8	25	*************************
1.0	27	***************************
1.2	21	**********************
1.4	13	*************
1.6	14	**************
1.8	15	***************
2.0	5	*****
2.2	4	****
2.4	2	**
2.6	1	*

Each * represents 2 observations

(c) Midpoint Count

0.0	2	*
0.2	6	***
0.4	16	********
0.6	31	****************
0.8	37	*******************
1.0	52	**************************
1.2	41	*********************
1.4	13	*******
1.6	2	*

Each * represents 5 observations

(d) Midpoint Count

−1.6	1	*
−1.2	0	
−0.8	0	
−0.4	13	***
0.0	62	**********
0.4	107	**********
0.8	17	****

Figure 5.21 Sampling distributions for MLE of (a) δ_1 , (b) δ_2 , (c) ln σ_ϵ and (d) ln σ_η for contaminated data

Two correlated explanatory variables

(a) Midpoint Count

−0.6	1	*
−0.4	2	**
−0.2	1	*
0.0	6	******
0.2	10	**********
0.4	12	**********
0.6	19	**********
0.8	26	**********
1.0	34	**********
1.2	31	**********
1.4	23	**********
1.6	19	**********
1.8	6	******
2.0	5	*****
2.2	4	****
2.4	1	*

(b) Midpoint Count

−0.2	2	**
0.0	5	*****
0.2	6	******
0.4	15	**********
0.6	31	**********
0.8	36	**********
1.0	33	**********
1.2	31	**********
1.4	24	**********
1.6	11	**********
1.8	4	****
2.0	1	*
2.2	1	*

Each * represents 2 observations

(c) Midpoint Count

−1.0	2	*
−0.8	1	*
−0.6	2	*
−0.4	5	***
−0.2	13	*******
0.0	13	*******
0.2	54	**********
0.4	66	**********
0.6	42	**********
0.8	2	*

Each * represents 2 observations

(d) Midpoint Count

−1.0	2	*
−0.8	1	*
−0.6	3	**
−0.4	6	***
−0.2	15	********
0.0	34	**********
0.2	60	**********
0.4	47	**********
0.6	23	**********
0.8	9	*****

Figure 5.22 Sampling distributions for robust estimator of (a) δ_1, (b) δ_2, (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for clean data

(a) Midpoint Count

−2.0	1	*
0.0	0	
0.2	9	*********
0.4	10	**********
0.6	33	**********
0.8	35	**********
1.0	43	**********
1.2	25	**********
1.4	23	**********
1.6	12	**********
1.8	6	******
2.0	3	***

Each * represents 2 observations

(b) Midpoint Count

0.2	10	*****
0.4	9	******
0.6	24	************
0.8	33	*****************
1.0	57	*****************************
1.2	33	*****************
1.4	26	*************
1.6	5	***
1.8	3	**

Each * represents 5 observations

(c) Midpoint Count

−0.4	1	*
−0.2	2	*
0.0	15	***
0.2	53	***********
0.4	104	**********************
0.6	24	*****
0.8	1	*

Each * represents 2 observations
1 observation below the first class

(d) Midpoint Count

−0.4	1	*
0.2	9	*****
0.0	22	***********
0.2	68	**********************************
0.4	77	**************************************
0.6	19	**********
0.8	3	**

Figure 5.23 Sampling distributions for MLE estimator of (a) δ_1, (b) δ_2, (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for clean data

(a) Midpoint Count

–0.4	3	***
–0.2	1	*
0.0	4	****
0.2	12	**********
0.4	14	**********
0.6	23	**********
0.8	23	**********
1.0	40	**********
1.2	34	**********
1.4	18	**********
1.6	12	**********
1.8	8	********
2.0	7	*******
2.2	0	
2.4	1	*

(b) Midpoint Count

0.0	4	****
0.2	13	**********
0.4	18	**********
0.6	21	**********
0.8	33	**********
1.0	36	**********
1.2	23	**********
1.4	30	**********
1.6	16	**********
1.8	4	****
2.0	2	**

Each * represents 2 observations

(c) Midpoint Count

–1.0	1	*
–0.8	0	
–0.6	0	
–0.4	3	**
–0.2	5	***
0.0	14	*******
0.2	34	**********
0.4	61	**********
0.6	66	**********
0.8	14	*******
1.0	2	*

(d) Midpoint Count

–1.2	1	*
–1.0	1	*
–0.8	1	*
–0.6	7	*******
–0.4	6	******
–0.2	15	***************
0.0	24	************************
0.2	47	**
0.4	41	***
0.6	36	************************************
0.8	15	***************
1.0	6	******

Figure 5.24 Sampling distributions for robust estimator of (a) δ_1 , (b) δ_2 , (c) ln σ_ϵ and (d) ln σ_η for contaminated data

(a) Midpoint Count

–1.2	1	*
–1.0	2	**
–0.8	0	
–0.6	3	***
–0.4	2	**
–0.2	3	***
0.0	6	******
0.2	8	********
0.4	15	***************
0.6	16	****************
0.8	28	****************************
1.0	38	**************************************
1.2	22	**********************
1.4	25	*************************
1.6	14	**************
1.8	6	******
2.0	4	****
2.2	1	*
2.4	5	*****
2.4	1	*

(b) Midpoint Count

–0.4	1	*
–0.2	1	*
0.0	4	****
0.2	12	**********
0.4	14	**********
0.6	23	**********
0.8	37	**********
1.0	38	**********
1.2	31	**********
1.4	17	**********
1.6	12	**********
1.8	6	******
2.0	4	****

Each * represents 2 observations

(c) Midpoint Count

0.0	1	*
0.2	7	****
0.4	16	********
0.6	30	**********
0.8	35	**********
1.0	51	**********
1.2	43	**********
1.4	14	*******
1.6	3	**

Each * represents 2 observations
5 observations below the first class

(d) Midpoint Count

–0.8	1	*
–0.6	2	*
–0.4	6	***
–0.2	12	******
0.0	29	**********
0.2	54	**********
0.4	63	**********
0.6	23	**********
0.8	5	***

Figure 5.25 Sampling distributions for MLE of (a) δ_1 , (b) δ_2 , (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for contaminated data

Special configurations of outliers

(a) Midpoint Count

−0.4	1	*
−0.2	1	*
0.0	8	********
0.2	6	******
0.4	17	*****************
0.6	23	**********************
0.8	21	*********************
1.0	34	**********************************
1.2	35	***********************************
1.4	20	********************
1.6	15	***************
1.8	11	***********
2.0	5	*****
2.2	3	***

(b) Midpoint Count

−0.2	2	**
0.0	4	**
0.2	7	*******
0.4	16	****************
0.6	26	**************************
0.8	37	*************************************
1.0	33	*********************************
1.2	29	*****************************
1.4	30	******************************
1.6	10	**********
1.8	5	*****
2.0	1	*

Each * represents 2 observations

(c) Midpoint Count

−1.8	1	*
−1.6	0	
−1.4	0	
−1.2	3	**
−1.0	0	
−0.8	2	*
−0.6	4	**
−0.4	3	**
−0.2	17	*********
0.0	29	***************
0.2	49	*************************
0.4	64	*********************************
0.6	26	**************
0.8	2	*

Each * represents 2 observations

(d) Midpoint Count

−1.0	1	*
−0.8	0	
−0.6	1	*
−0.4	3	**
−0.2	7	****
0.0	19	**********
0.2	46	**********
0.4	57	**********
0.6	44	**********
0.8	21	**********
1.0	1	*

Figure 5.26 Sampling distributions for robust estimator of (a) δ_1, (b) δ_2, (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for contamination at first observation

Each * represents 2 observations

(a) Midpoint Count

−2.5	1	*
−2.0	0	
−1.5	4	**
−1.0	6	***
−0.5	14	*******
0.0	19	**********
0.5	44	**********
1.0	54	**********
1.5	37	**********
2.0	14	*******
2.5	2	*
3.0	2	*
3.5	2	*
4.0	1	*

(b) Midpoint Count

−0.4	1	*
−0.2	1	*
0.0	3	***
0.2	3	***
0.4	12	**********
0.6	23	**********
0.8	27	**********
1.0	29	**********
1.2	37	**********
1.4	28	**********
1.6	12	**********
1.8	7	*******
2.0	1	*
2.2	1	*
2.4	1	*

Each * represents 2 observations

(c) Midpoint Count

−14.0	43	**********
−12.0	1	*
−10.0	1	*
−8.0	0	
−6.0	0	
−4.0	0	
−2.0	0	
0.0	88	**********
2.0	67	**********

Each * represents 5 observations

(d) Midpoint Count

−14.0	7	**
−12.0	0	
−10.0	0	
−8.0	0	
−6.0	0	
−4.0	0	
−2.0	1	*
0.0	89	**********
2.0	102	**********
4.0	1	*

Figure 5.27 Sampling distributions for MLE of (a) δ_1, (b) δ_2, (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for contamination at first observation

(a) Midpoint Count

−0.4	2	**
−0.2	2	**
0.0	6	******
0.2	9	*********
0.4	13	**********
0.6	17	**********
0.8	34	**********
1.0	27	**********
1.2	32	**********
1.4	30	**********
1.6	14	**********
1.8	6	******
2.0	4	****
2.2	3	***
2.4	1	*

(b) Midpoint Count

−0.4	1	*
−0.2	0	
0.0	7	*******
0.2	7	*******
0.4	12	************
0.6	27	***************************
0.8	41	***
1.0	28	****************************
1.2	31	*******************************
1.4	28	****************************
1.6	12	************
1.8	4	****
2.0	1	*
2.2	1	*

Each * represents 2 observations

(c) Midpoint Count

−1.2	1	*
−1.0	1	*
−0.8	0	
−0.6	2	*
−0.4	5	***
−0.2	9	*****
0.0	15	********
0.2	51	***************************
0.4	70	**********************************
0.6	44	**********************
0.8	2	*

Each * represents 2 observations

(d) Midpoint Count

−1.0	1	*
−0.8	2	*
−0.6	3	**
−0.4	5	***
−0.2	15	********
0.0	35	******************
0.2	51	***************************
0.4	47	************************
0.6	31	****************
0.8	10	*****

Figure 5.28 Sampling distributions for robust estimator of (a) δ_1, (b) δ_2, (c) $\ln \sigma_\epsilon$ and (d) $\ln \sigma_\eta$ for contamination at last observation

Each * represents 2 observations

(a) Midpoint Count

−2.0	4	**
0.0	5	***
2.0	27	**********
4.0	45	**********
6.0	65	**********
8.0	36	**********
10.0	11	******
12.0	4	**
14.0	3	**

(b) Midpoint Count

−5.0	1	*
−4.0	7	*******
−3.0	25	**********
−2.0	45	**********
−1.0	21	**********
0.0	13	**********
1.0	21	**********
2.0	20	**********
3.0	20	**********
4.0	16	**********
5.0	7	*******
6.0	3	***
7.0	1	*

Each * represents 2 observations

(c) Midpoint Count

−14.0	97	**********
−12.0	2	*
−10.0	3	**
−8.0	0	
−6.0	0	
−4.0	0	
−2.0	0	
0.0	7	****
2.0	34	**********
4.0	57	**********

Each * represents 2 observations

(d) Midpoint Count

−14.0	14	*******
−12.0	0	
−10.0	0	
−8.0	0	
−6.0	0	
−4.0	0	
−2.0	0	
0.0	20	**********
2.0	79	**********
4.0	87	**********

Figure 5.29 Sampling distributions for MLE of (a) δ_1 , (b) δ_2 , (c) ln σ_ϵ and (d) ln σ_η for contamination at last observation

Regression plots from GLST procedure using whole data set

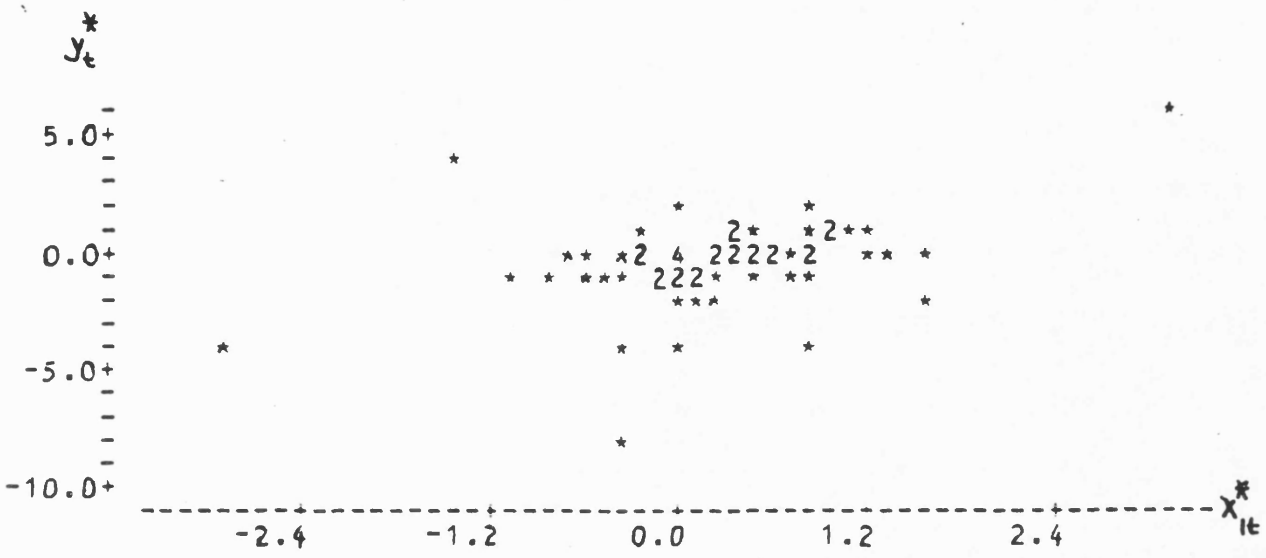


Figure 5.30a Plot of y_t^* against x_{1t}^*

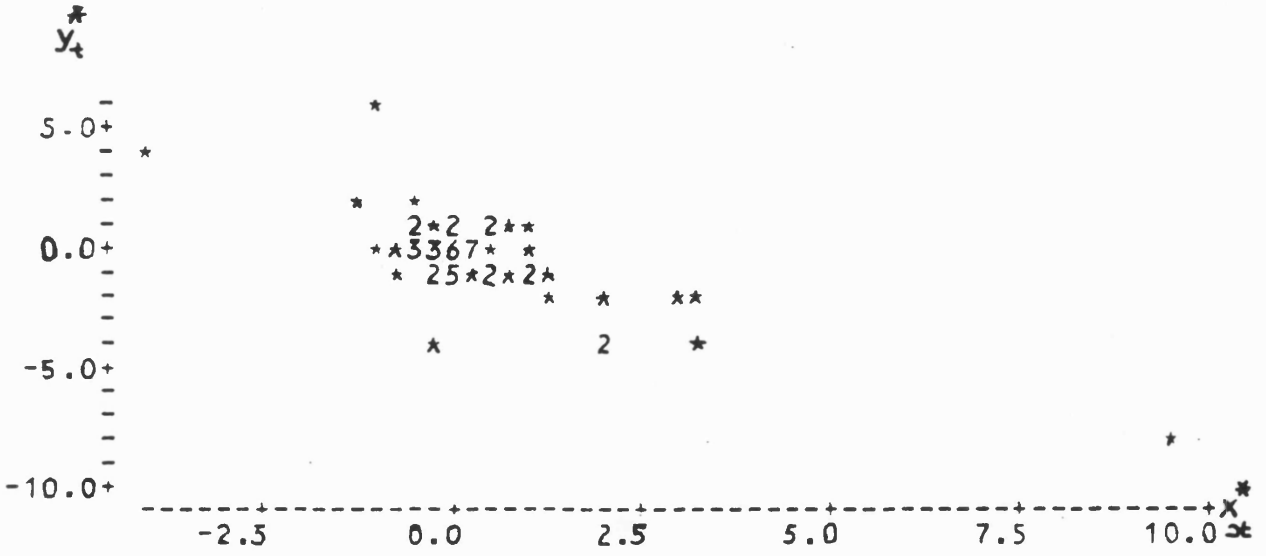


Figure 5.30b Plot of y_t^* against x_{2t}^*

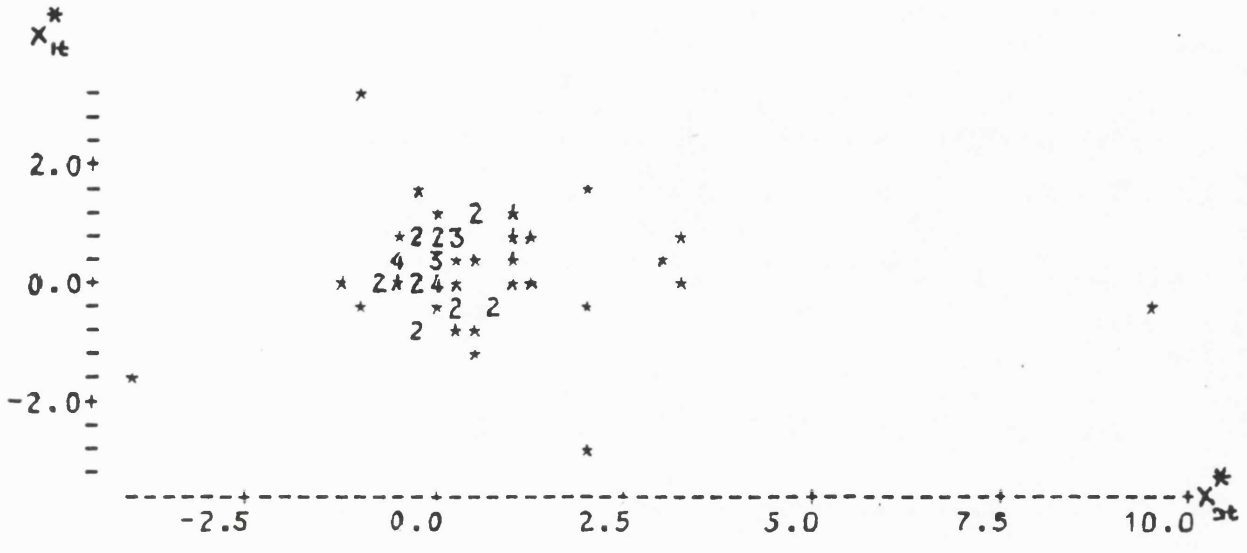


Figure 5.30c Plot of x^*_{1t} against x^*_{2t}

Regression plots from GLST procedure using trimmed data set

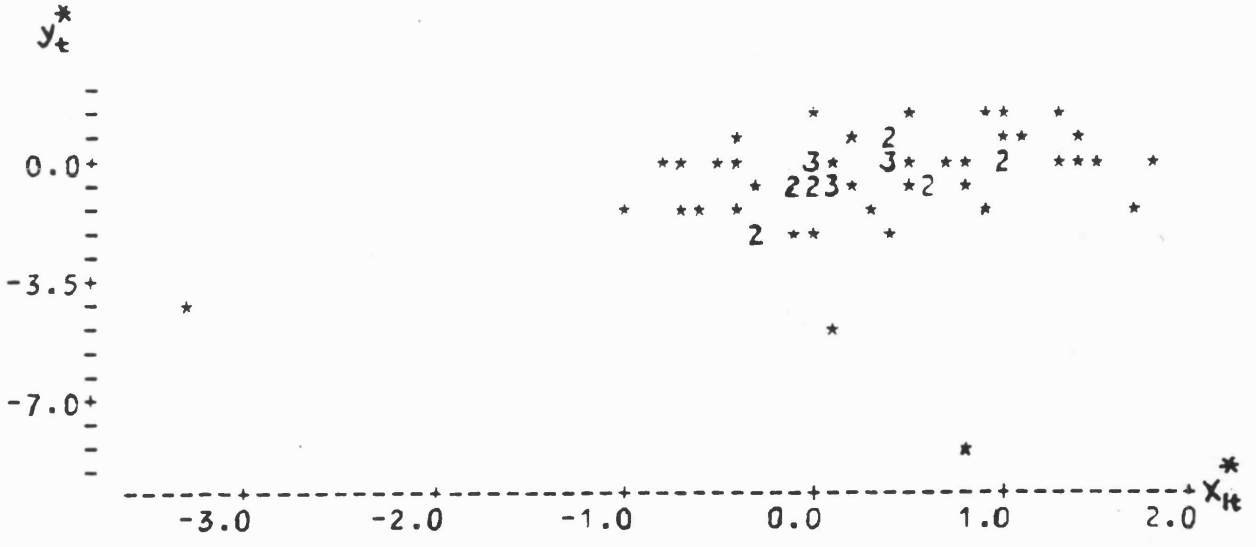


Figure 5.31a Plot of y^*_t against x^*_{1t}

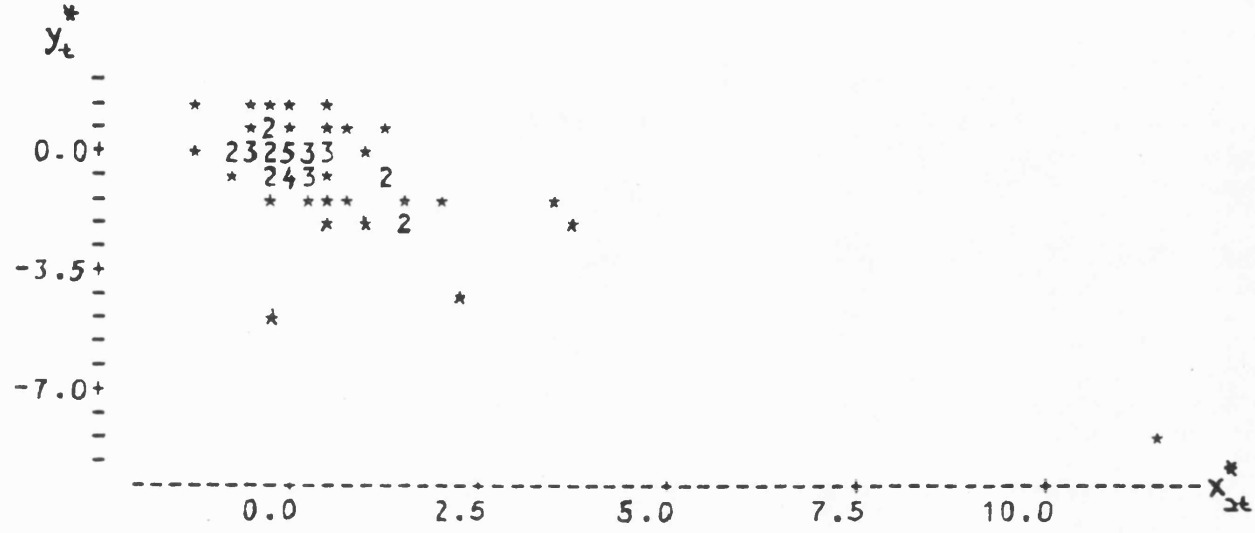


Figure 5.31b Plot of y_t^* against x_{2t}^*

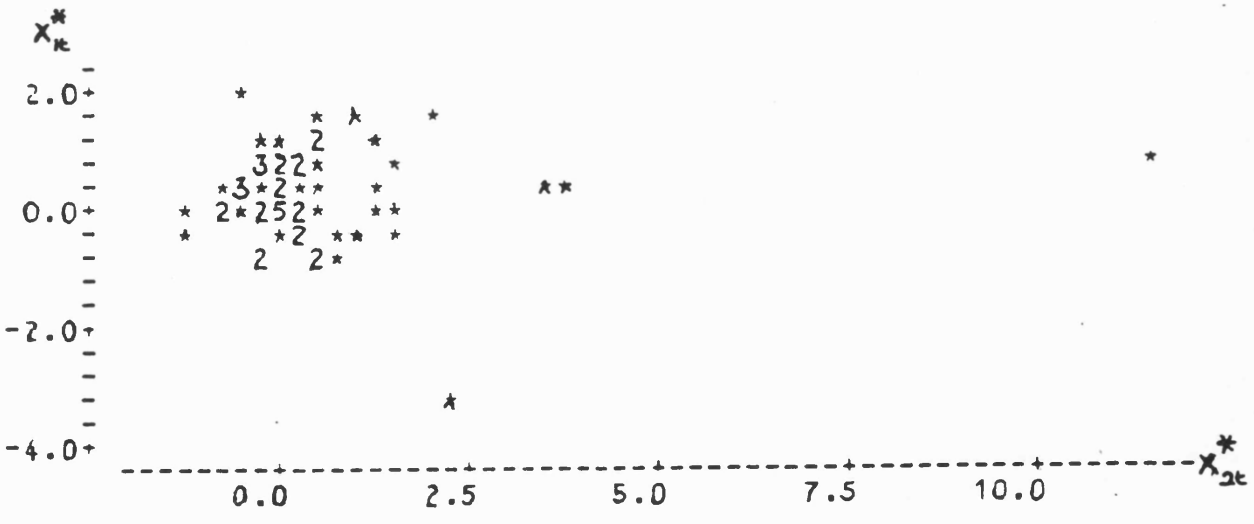


Figure 5.31c Plot of x_{1t}^* against x_{2t}^*

Regression plots from robustified GLST procedure

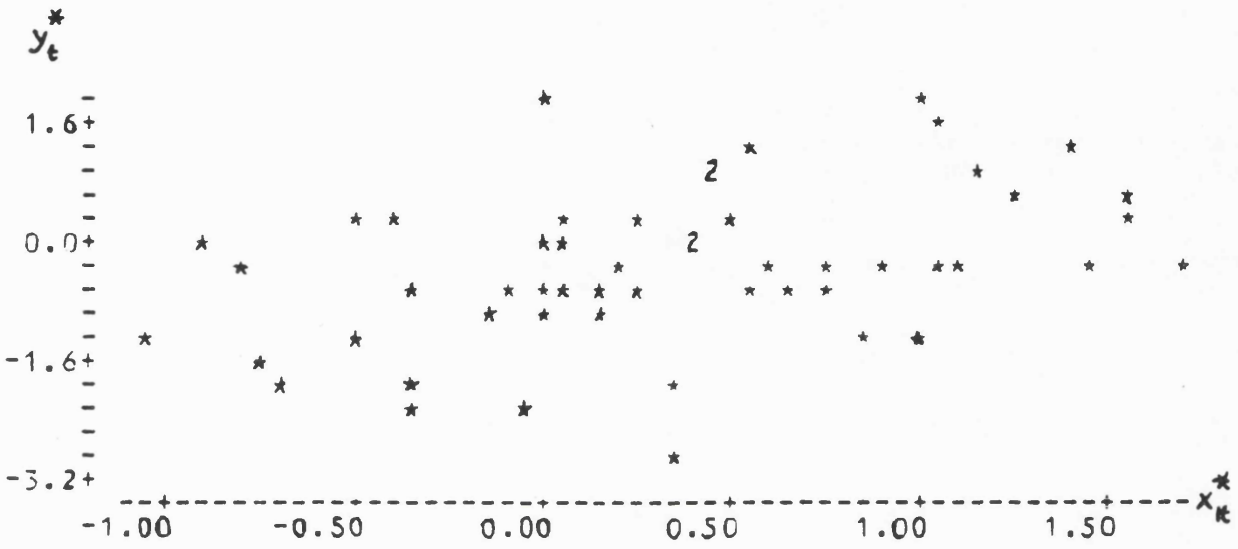


Figure 5.32a Plot of y^*_t against x^*_{1t}

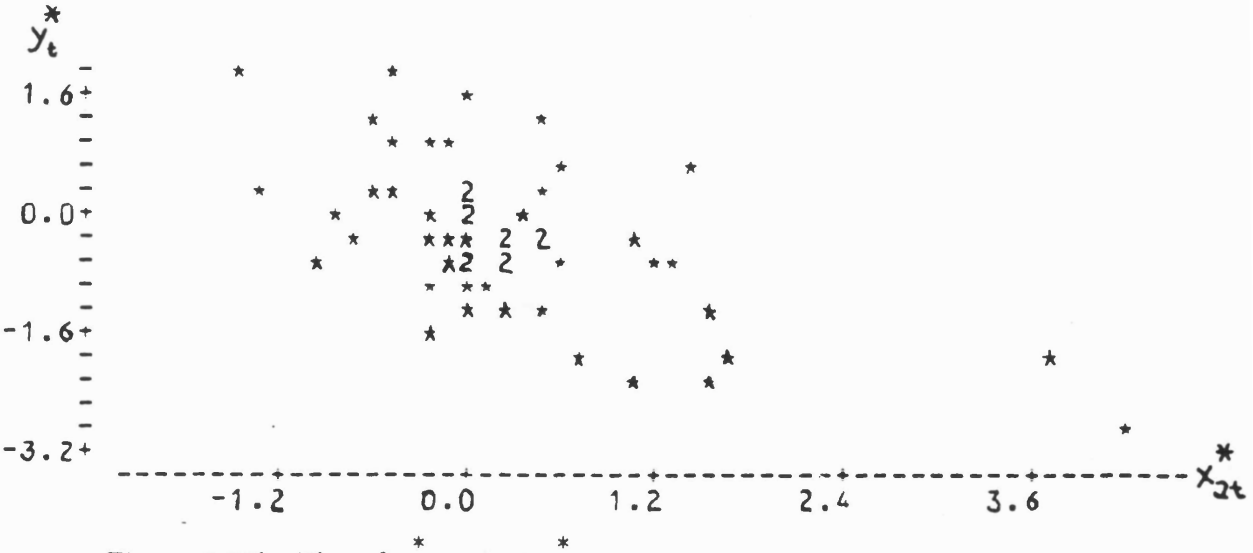


Figure 5.32b Plot of y^*_t against x^*_{2t}

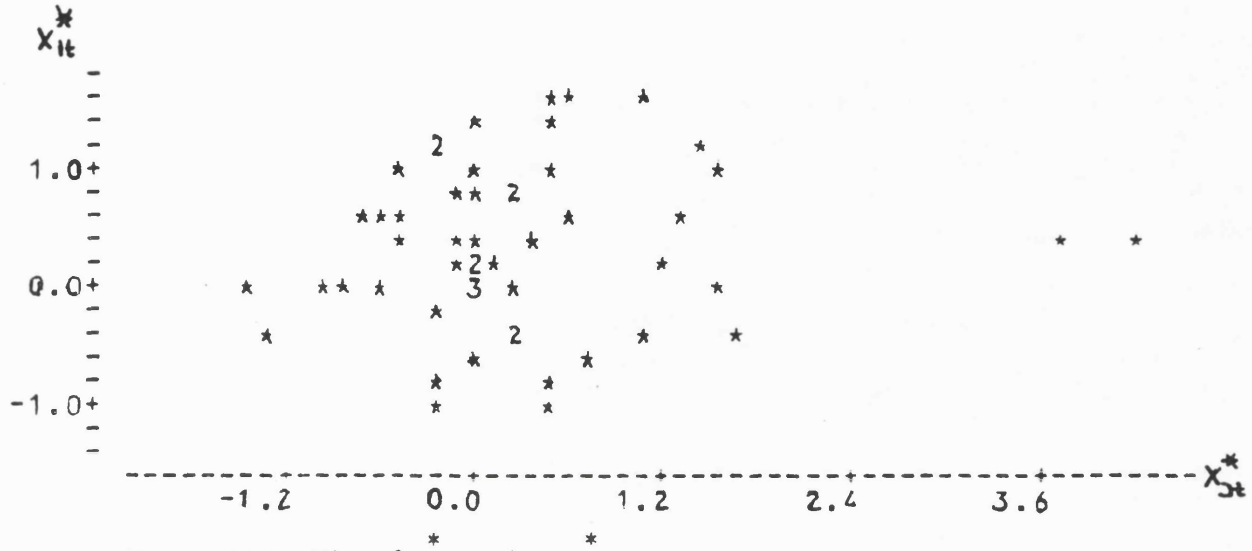


Figure 5.32c Plot of x_{1t}^* against x_{2t}^*

Bibliography

Abraham,B. and Box,G.E.P. (1979)
"Bayesian Analysis of some Outlier Problem in Time Series"
Biometrika,66,2,229–236

Anderson,B.D.O. and Moore,J.D. (1979)
Optimal Filtering
Englewood Cliffs : Prentice–Hall

Akaike,H. and Ishiguro,H. (1980)
"Trend Estimation with Missing Observations"
Ann. of the Inst. of Stat. Maths.,32,b,481–488

Alspach,D.L. & Sorenson,H.W. (1971)
"Recursive Bayesian Estimation using Gaussian Sums"
Automatica,vol. 6,465–479

Ansley,C.F. and Kohn,R. (1985)
"Efficient estimation and prediction in time series regression models"
Biometrika,72,3,694–7

Ansley,C.F. and Kohn,R. (1989)
"A fast algorithm for signal extraction , influence and cross–validation in state space models"
Biometrika,76,1,66–79

Atkinson,A.C. (1986)

Plots , Transformations and Regression – An introduction to Graphical Methods of Diagnostics Regression Analysis

Oxford : Clarendon Press

Atkinson,A.C. (1986)

"Masking Unmasked"

Biometrika,73,533–541

Barnett,V. and Lewis,T. (1977)

Outliers in Statistical Data

New York : John Wiley publ.

Basawa,I.V. , Huggins,R.M. and Stuadte,R.G. (1985)

"Robust Test for Time Series with an Application to First–order Autoregressive Process"

Biometrika,72,559–571

Box,G.E.P. and Jenkins,G.M.M. (1976)

Time Series Analysis : forcasting and control

Holden–day

Bruce,A.G. and Martin,R.D. (1989)

"Leave–k–out diagnostics for time series"

JRSS B,51,363–424

Chu,K. (1973)
"Estimation and Decision for Linear Systems With Elliptical Random Processes"
IEEE Trans. Auto. Control,vol AC–20,107–110

Cook,D. (1986)
"Assesment of Local Influence"
JRSS B,48,133–169

Cook,D. and Weisberg,S. (1982)
Residuals and Influence in Regession
New York : Chapman anf Hall

Denby,L. and Martin,R.D. (1979)
"Robust Estimation on the First Order Autoregressive Parameter"
JASA,74,140–146

Dejongh,P.J. and Dewet,T. (1985)
"Trimmed mean and bounded influence estimators for the parameters of the AR(1) process"
Commun. Statist.–Theor. Meth.,14(6),1361–1375

Ershov,A.A. and Liptser,R.S. (1978)
"Robust Kalman Filter and Discrete Time"
Automation and Remote Control,360–369

Durbin, J and Watson, G.S. (1951)
"Testing for serial correlation in least squares regression II"
Biometrika , 38 , 159 -78

Everitt,B.S. and Hand,D.J. (1981)

Finite Mixture Distributions

New York : Chapman and Hall

Fox,A.J. (1972)

"Outliers in Time Series"

JRSS B,34,350–363

Gnanadesikan,R. (1977)

Methods for Statistical Data Analysis of Multivariate Observations

New York : John Wiley

Guttman,I. and Pena,D. (1985)

Comment on "Dynamic Generalisation Linear Models and Bayesian Forcasting" by West,M. , Harrison,P.J. and Migon,H.S.

JASA,74,140–146

Hampel,F.R. (1971)

"A General Qualitative Definition of Robustness"

Ann. of Maths. Stats.,42,1887–1896

Hampel,F.R. (1974)

"The Influence Curve and Its Role in Robust Estimation"

JASA,69,383–393

Hampel,F.R. , Ronchetti,E.M. , Rouseeuw,P.J. and Stahel,W.A. (1986)
Robust Statistics : The Approach Based on Influence Functions
John Wiley and Sons publ.

Harrison,P.J. and Stevens,C.F. (1976)
"Bayesian forcasting"
JRSS B,34,1–41

Harvey,A.C. (1981)
Time Series Models
Philip Allan publ.

Harvey,A.C. (1982)
"Estimating Procedures for a Class of Univariate Time Series Models"
L.S.E. Econometrics Programme,Discussion Paper no. A28

Harvey,A.C. (1982)
"An Alternative Framework for Time Series Model Building and its Implications for Econometrics"
L.S.E. Econometrics Programme,Discussion Paper no. 32

Harvey,A.C. (1989)
Forecasting , structural time series models and the Kalman filter
Cambridge University Press

Harvey,A.C. and Durbin,J. (1985)
"The Effects of Seat Belt Legislation on British Road Casualties : A Case Study in Structural Time Series Modelling"
JRSS A,149,187–227

Harvey,A.C. and Phillips,G.D.A. (1979)
"Maximum likelihood estimation of regression models with autoregressive–moving average disturbances"
Biometrika,66,1,49–58

Harvey,A.C. and Peters,S. (1984)
"Estimating Procedures for Structural Time Series Models"
L.S.E. Econometrics Programme,Discussion Paper no. A44

Hoaglin , Mosteller and Tukey (1983)
Understanding Robust and Exploratory Data Analysis
John Wiley & Sons publ.

Huber,P.J. (1964)
"Robust Estimation of a Location Parameter"
Ann. Math. Statist.,vol 35,1,73–101

Huber,P.J. (1972)
"The 1972 Wald Lecture–Robust Statistics:A Review"
Ann. Math. Statist.,vol 43,4,1041–1067

Huber,P.J. (1973)
"Robust Regression:Aysmptotics,Conjectures and Monte Carlo"
The Ann. of Stats,1,799–821

Huber,P.J. (1977)
Robust Statistical Procedures
Philadelphia : SIAM

Kitagawa,G. (1979)
"Robust Estimation Through the Modelling of Data Generation"
Research Memorandom,177,The Institute of Statistical Maths.,Toyko

Kitagawa,G. (1987)
"Non–Gaussian State Space Modelling of Nonstationary Time Series" (with discussion)
JASA,vol. 82,1032–1063

Koenker,R. and Bassett,G. (1978)
"Regression Quantiles"
Econometrica,46,1,33–49

Koenker,R. and D'Orey (1987)
"Computing Regression Quantiles"
Applied Statistics,36,3,383–393

Kunsh,H. (1984)
"Infinitesimal Robustness for Autoregressive Processes"
Ann. of Stats,12,843–863

Ljung,L. (1978)
"Convergence Analysis of Parametric Identification Methods"
IEEE Trans. Aut. Contr.,vol AC–23,770–783

Martin,R.D. (1980)
"Robust Estimation In Autoregressive Models"
Directions in Time Series,228–254

Martin,R.D. & Masreliez,C.J. (1975)
"Robust Estimation via Stochastic Approximation"
IEEE Trans. Inform. Theory,vol IT–21,263–271

Martin,R.D. & Masreliez,C.J. (1977)
"Robust Bayesian Estimation for the Linear Model & Robustifying the Kalman Filter"
IEEE Trans. Auto. Control,vol AC–22,361–371

Martin,R.D. , Samorov,A. and Vandaele,W. (1983)
"Robust Methods for ARIMA Models"
Applied Time Series Analysis of Economic Data,153–167

Martin,R.D. and Yohai,V.J. (1985)
"Robustness in Time Series and Estimating ARIMA Models"
Handbook of Stats.,5,119–155

Martin,R.D. and Yohai,V.J. (1986)
"Influence Functional for Time Series"
Ann. of Stats.,14,781–855

Masreliez,C.J. (1975)
"Approximate Non–Gaussian Filtering with Linear State & Observation relation"
IEEE Trans. Auto. Control,vol AC–20,107–110

Meinhold,R.J. and Singpurwalla,N.D. (1989)
"Robustification of Kalman filter models"
JASA,84,479–486

Ruppert,D. and Carroll,R.J. (1980)
"Trimmed least squares estimation in the linear model"
JASA,75,823–838

Stigler,S.M. (1977)
"Do Robust Estimators Work with Real Data?"
Ann. of Stats.,5,1055–1098

Stoodley,D.C. & Mirinia,M. (1979)

"The Automatic Detection of Transients , Step Changes & Slope Changes in the Monitoring of Medical Time Series"

The Statistician,vol 28,163–170

Welsh,A.H. (1987)

"The trimmed mean in linear model"

The Ann. of Stats.,15,20–36

West,M. (1981)

"Robust Sequential Approximate Baysesian Estimation"

JRSS B,43,157–166